ROMANIAN NOMINAL INFLECTION: A REALIZATIONAL APPROACH

ALAIN KIHM

Résumé. Le roumain est la seule langue romane vivante dont les noms et les adjectifs (les nominaux) se fléchissent pour le cas. Cette flexion se résume à l'opposition de deux cas: cas direct, nominatif et accusatif, et cas oblique, génitif et datif. Les nominaux masculins ne la manifestent que sur les déterminants: cf. *un prieten* 'un ami' vs. *unui prieten* 'de/à un ami'. Seuls les nominaux féminins fléchissent leur base: cf. *o pisică* 'un chat' vs. *unei pisici* 'de/à un chat'. La flexion féminine présente en outre une particularité remarquable: la forme du cas oblique singulier est identique au pluriel, où le cas n'est distingué que sur le déterminant: cf. *nişte pisici* 'des chats' vs. *unor pisici* 'des/à des chats'. Le présent article propose un explication synchronique de ce phénomène dans le cadre de la morphologie réalisationnelle: on fait l'hypothèse que l'unique exposant flexionnel du roumain possède une propriété « bascule » (*toggle*) telle qu'il exprime le cas non-défaut (oblique) quand le nombre a sa valeur par défaut (singulier); sinon, il exprime la valeur non-défaut du nombre (pluriel). On montre en outre pourquoi les deux valeurs non-défaut du cas et du nombre ne peuvent se cumuler, et pourquoi la flexion de la base est limité aux seuls nominaux « tout féminins », en sorte que les prétendus neutres, en réalité ambigènes, ne la présentent pas. La flexion nominale roumaine apparaît ainsi analogue dans son système à celle de l'ancien français.

1. INTRODUCTION

The present paper proposes a new account of Romanian nominal inflection. In addition to being assigned to a given gender or word-class Romanian nouns and adjectives inflect for number and, less typically for a Romance language, for case. Case inflection, however, is extremely sparse, as it involves one exponent in feminine nominals only. I will therefore argue that such a system cannot be handled with the usual devices for accounting for homonymous cells in paradigms, such as rules of referral and the notion of syncretism (Stump 2001, Chapter 7; also see Academia Română (AR) 2005: 70ff.). Indeed, syncretism fails to account for the peculiar marking pattern of the lone functional exponent of feminine nouns and adjectives, namely that it marks non-default case (i.e. oblique) when number is default (i.e. singular), whereas it marks only number when the latter is non-default (i.e., plural). I therefore propose that a new type of exponent, provided with what I call a 'toggle' property, ought to be recognized.

RRL, **LII**, *3*, p. 255–302, Bucureşti, 2007

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

Toggle exponents are **unique**, that is neither syncretic nor homonymous, exponents that express either one of two non-default feature values. Their assumption is part of a model of the interface between lexical matrices and phonological realizations ('words' and 'word-forms' in the sense of Matthews 1972), consisting in what I call Concrete Lexical Representations (CLRs). CLRs describe morphological words, neither words nor word-forms, and they constitute subsets of lexical matrices including only those elements of the matrix that play a morphological role, directly by being associated with an exponent and/or indirectly through the processes they trigger along the syntactic chain. In this way CLRs may be viewed as representational, set-theoretic alternatives to realization rules within a globally realizational theory of morphology (see Stump 2001).

The paper proceeds as follows. First, I give a brief description of Romanian nominal inflection (section 2) and I put forward the empirical and theoretical issues that it raises (section 3). In section 4 I present the framework I propose for dealing with these issues. A formal characterization of the interface between lexical matrices and phonological realization is attempted (section 5). Then comes a formal analysis in the framework presented in sections 4 and 5 of the three grammatical categories for which Romanian nouns and adjectives are marked or inflect: gender, number, and case (sections 6, 7, and 8). Finally the results are summed up (section 9) and some generalizations are attempted through comparison with other Romance languages and beyond (section 10).

2. THE FACTS WITH ROMANIAN NOMINAL INFLECTION

Romanian is well-known for being the only modern Romance language that kept something of Latin case inflection outside personal pronouns.[1] According to traditional grammars Romanian nouns and adjectives, inflect for five cases : nominative, genitive, dative, accusative, and vocative, with more or less the same meanings and uses as in Latin, and they belong to three genders : masculine, feminine, and neuter, thus making Romanian the most conservative of Romance languages, in this domain at least (see, e.g., *Gramatica limbii române* (GA), vol. I 1963 : 57ff., 74ff.; Daniliuc & Daniliuc 2000: XIII). An unprejudiced look at the actual paradigms gives a rather different picture, however.

Take gender first. Neuter is certainly not a morphological class in Romanian as it is in Latin (cf. *templ**um*** 'temple', *mar**e*** 'sea', *corn**u*** 'horn') or the Slavic languages (cf. Bulgarian *sel**o*** 'village', *det**e*** 'child'). So-called neuter nouns are indeed those nouns which are masculine in the singular and feminine in the plural as shown by the form of the enclitic definite determiner or article: e.g., *trenul* /

[1] Klausenburger (2001: 119) insists that Romanian does not directly continue Latin in this area. It is true to a large extent, but not entirely as we shall see. See Mallinson (1984, 1988) for a balanced presentation.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

trenurile ‘the train(s)’, in which *-(u)l* and *-le* are the masculine singular and feminine plural forms of the article respectively. They should therefore more properly be called ‘ambigeneric’ (see Lombard, Gâdei 1981), and they accordingly control masculine agreement of attribute and predicate adjectives in the singular, but feminine agreement in the plural, as illustrated below with the reputedly neuter noun *teatru* ‘theatre’:[2]

(1a) teatr-ul frumos
theatre-the$_{M.SG}$ beautiful$_{M.SG}$
the beautiful theatre
(1b) Teatr-ul e frumos.
theatre-the$_{M.SG}$ is beautiful$_{M.SG}$
The theatre is beautiful
(2a) teatre-le frumoase
theatres-the$_{F.PL}$ beautiful$_{F.PL}$
the beautiful theatres
(2b)Teatre-le sunt frumoase.
theatres-the$_{F.PL}$ are beautiful$_{F.PL}$
The theatres are beautiful.

Even the plural ending /-urĭ/ (e.g., *trenul* / *trenurile* ‘the train(s)’) is by no means a reliable clue : not all ‘neuters’ take it (cf. *teatru* / *teatre*); a number of feminines have it (e.g., *gheață* / *ghețuri* ‘ice / icicles’, *lipsă* / *lipsuri* ‘shortage(s)’, *marfă* / *mărfuri* ‘merchandise(s)’, *treabă* / *treburi* ‘work(s)’, etc. – cf. GA I : 69).

On the semantic side the Romanian ‘neuter’ has no hard-and-fast classifying value either – like, say, the English neuter. True, a great majority of the nouns assigned to it refer to inanimate entities. A small number of items even present variation in gender according to whether the reference is to animate or inanimate: cf. *cap* / *capete* ‘head(s)’ ‘neuter’ vs. *cap* / *capi* ‘chief(s)’ masculine, or *zmeu* masculine in the sense of ‘dragon’, but ‘neuter’ in the sense of ‘kite’. Yet, it is also a fact that (a) all nouns with inanimate reference are not, far from it, neuter or ambigeneric (cf. *colț*$_{masc.}$ ‘corner’, *casă*$_{fĕm.}$ ‘house’) ; (b) a significant and apparently growing (Pană Dindelegan, p.c.) number of nouns refering to animate, including human, beings are ambigeneric (e.g., *animal* ‘animal’, *personaj* ‘character’, *planton* ‘orderly’, *tist* ‘officer’, *star* ‘(movie) star’, etc. – see Lombard & Gâdei 1981: 10-11). Interestingly, almost all these ambigeneric animates are borrowings from French or, increasingly, English. There may be a tendency here : ambigenericity as a class for new borrowings. However that may be, though, it imparts no ‘natural’ classificatory value to the Romanian ‘neuter’. Neither from a

[2] In order to optimize the alignment of the Romanian data with the glosses, I separate the article from the stem with a hyphen, not present in Romanian orthography.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

semantic nor from a morphological viewpoint can it be regarded as a 'third' gender as implied by most traditional and recent treatments (see, e.g. Farkas 1990; Lumsden 1992).

Let us turn to case.[3] Compared to Latin, Romanian presents us with an impoverished inflection that overtly contrasts two cases only: nominative-accusative and genitive-dative, which I relabel direct and oblique case respectively. Moreover, not all nouns manifest this contrast. Only feminine nouns do, as shown by the following paradigm for *casă* 'house':

(3)

	Singular	Plural
Direct	cas-ă	cas-e
Oblique	cas-e	cas-e

Feminine nouns thus present a four-cell paradigm that includes only one functional exponent, namely *-e*.[4] The *-ă* ending of the singular direct form doesn't count as a functional exponent, as it does not express a morphosyntactic feature such as case or number, but it is a word-class or gender marker (see below for details).

Setting up such a paradigm for masculine nouns would be otiose, since both cells in the singular and the plural would always be identically filled. I therefore posit a two-cell paradigm where only number is distinguished, as in (4) for *prieten* 'friend':

(4)

Singular	Plural
prieten	prieten-i

One thus observes complete identity of putative case forms for unarticulated singular and plural masculine and plural feminine nouns and adjectives.[5] (No mention of 'neuter' is called for since, as just seen, it results from the conjunction of masculine singular with feminine plural.) In other words, all 'inflected' forms in such nouns turn out to be identical to the direct form, which is itself either identical to the root (cf. masculine *prieten* 'friend', neuter *toc* 'pen') or consists in the root plus a final vowel (cf. masculine *fiu* 'son' or *câine* 'dog', neuter *teatru* 'theater') or a plural ending (cf. *prieteni* 'friends', *tocuri* 'pens'). Case distinctions are then uniquely expressed on the functional modifiers as the following example shows :

[3] Of the putative five cases in the paradigm, I leave aside the only nonstructural one, that is the vocative, as its use is both quite limited and subject to complex and rapidly changing pragmatic and sociolinguistic constraints (for a careful study of which see Croitor Balaciu 2004). Moreover, I deal with nouns and adjectives only. Pronouns and various kinds of determiners, more richly inflected all of them, are only considered for contrastive purposes.

[4] The exponent is *-i* in a different subclass of feminine nouns (see below).

[5] Following tradition I call 'unarticulated' (*nearticulat*) nouns which do not bear the suffixed definite article. Unarticulated nouns can be indefinite (*un prieten* 'a friend') or definite (*acest prieten* 'this friend'). Articulated nouns, in contrast, are always definite.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

(5) I-am arătat acest-ui profesor ruine-le un-ui vechi oraş
to.him-I.have shown this-OBL teacher ruins-the$_{F.PL.DIR}$ a-OBL old town
I showed this teacher the ruins of an old town

Similar to *acest* 'this' and *un* 'a', one finds *fiecare* 'each', *alt* 'other', and so forth. Note that genitive and dative are again identical in both numbers, forming what I call the oblique case (cf. *unui* 'of/to a', *unor* 'of/to some', *acestui* 'of/to this', *acestor* 'of/to these', *fiecărui* 'of/to each', *altui* 'of/to another', *altor* 'of/to other') and that accusative is nondistinct from nominative, both forming the direct case. Case contrasts, wherever they show up, are thus always binary rather than fourfold.

In a similar way, case distinctions for non-feminine-singular articulated nouns appear exclusively on the article as illustrated below:[6]

(6) I-am arătat profesor-u-lui ruine-le vechi-u-lui
to.him-I.have shown teacher-EV-the$_{M.SG.OBL}$ ruins-the$_{F.PL.DIR}$ old-FV-the$_{M.SG.OBL}$
oraş.
town
I showed the teacher the ruins of the old town.

Here too genitive and dative as well as nominative and accusative are identical in both numbers: cf. *profesorul* 'the teacher', subject or object; *profesorului* 'of/to the teacher'; *profesorii* 'the teachers'; *profesorilor* 'of/to the teachers'. In *profesorii* the first /i/ is the plural ending (cf. *profesori* 'some teachers'), and the second /i/ is the masculine plural article. Both are pronounced as one /i/ – whereas in *profesori* final plural /i/ is the so-called 'asyllabic' [ĭ] which surfaces (mainly) as a palatal feature on the preceding consonant.

Only with feminine nouns and adjectives, therefore, do we observe an internal difference in case marking, and an intriguing one at that : singular feminine nouns and adjectives in the oblique case, articulated or unarticulated, take the form of the **plural**. This is apparent in (3) and it is especially striking when pluralization entails a metaphonical (umlaut) process affecting the root vowel as in *fată* / *fete* 'girl(s)':[7]

(7) rochi-a un-ei fet-e
dress- the$_{F.SG.DIR}$ a-F.SG.OBL girl-F.SG.OBL
the dress of a girl
(8) rochi-a fet-e-i

[6] In adjective-noun or noun-adjective phrases the enclitic article appears only once on the first term. For noun phrases and genitive constructions in Romanian, see Cornilescu (1992), Dobrovie-Sorin (2000). EV in (6) means 'epenthetic vowel'.

[7] I will assume this process belongs in phonology and not concern myself with it. This is an oversimplification, innocuous for my purposes.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

dress-the$_{F.SG.DIR}$ girl-F.SG.OBL-the$_{F.SG.OBL}$
the dress of the girl

In (7) the singular unarticulated oblique case appears fully identical with the unarticulated plural. Compare (9) and (10):

(9) Le cunosc pe acest-e fet-e
them$_F$ I.know PE these-F.PL.DIR girl-F.PL
I know these girls.[8]
(10) Cunosc nume-le acest-or fet-e
I.know name-the$_{M.SG.DIR}$ these-PL.OBL girl-F.PL
I know the name of these girls.[9]

The following table may help clarify such complex data :

(11)

	Unarticulated	Articulated
Singular	DIR fată OBL fete	fat-a fete-i
Plural	DIR fete OBL fete	fete-le fete-lor

The whole puzzle of the Romanian nominal inflection lies there. How can we account for this surprising 'number polarity' of feminine nominals, to paraphrase the 'gender polarity' of some Afroasiatic languages? And is it related to the apparently real gender polarity of neuter or ambigeneric nouns?

Before we can tackle these questions, however, we must first get a clear idea of the theoretical issues behind the facts just sketched. Then it is necessary to outline the formal framework to be used in order to deal with these issues. That is what I shall now be doing in the following three sections.

3. THE ISSUES WITH ROMANIAN NOMINAL INFLECTION

To begin with, there should be no doubt left that, far from comprising four cases (vocative aside), nominal inflection in Romanian actually consists in the binary contrast of a morphologically unexpressed default direct case and a morphologically expressed non-default oblique case. The former is the case of the (nominative) subject of the predicate and of the direct (accusative) object of verbs

[8] Human direct objects such as *fete* 'girls' must be governed by the preposition *pe*, and they must be resumed by a proclitic coindexed pronoun, here *le* 'them (feminine)'.

[9] The *-le* in *nume-le* is an allomorph of *-(u)l*, not to be confounded with feminine plural *-le*.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

and of most prepositions (cf. *cu o fată* 'with a girl'); the latter is the case of the genitive complement of nouns (cf. [7]) and of the indirect (dative) object of verbs (cf. [6]). Moreover, as we saw in the preceding section, overt contrast is a property of feminine singular nouns and adjectives only (cf. [3] and [11]).

Given this, three issues arise: (i) Why is it that only feminine singular nouns and adjectives overtly inflect for case? (ii) Does it entail that only they include a case feature, or do masculine and/or plural nouns and adjectives also include this feature, only not overtly? (iii) How do we account for the identity of the case exponent with the number exponent? I will concern myself with the latter issue in this section, as it is immediately crucial for selecting the proper theoretical framework for dealing with the whole phenomenon; and because answering the first two questions requires analyses that can only be achieved once the framework issue is settled.

Two accounts that may be thought of fall short of the expected result. One, homonymy, is traditional. According to it, Romanian grammar includes two phonologically identical morphemes associated to the distinct feature sets {GDR:fem, NUM:sg, CASE:obl} and {GDR:fem, NUM:pl, CASE:null}.[10] Phonological identity is regarded as an accident to be explained by history, namely the fact that most nouns like *fată* come from Latin 1st declension nouns, in which the genitive-dative singular and nominative plural endings look alike as well: cf. *rosae* /ros-ae$_{GEN/DAT.SG}$/ 'of/to the/a rose' vs. *rosae* /ros-ae$_{NOM.PL}$/ '(the) roses'.

Unobjectionable as this account is, it suffers from two severe drawbacks. One is that it is irrelevant from the perspective of **synchronic** Romanian grammar defined as the implicit knowledge of speakers-hearers, most of whom know nothing of the history of their language, while the erudition of those who do know something does not belong to linguistic competence, but to a different kind of knowledge. In other words, learning the history of the phenomenon does not dispense us from trying to devise a plausible model of the present speakers' competence in this area.

The second shortcoming of the homonymy account is that it makes nothing of the obvious observation that, discounting the identical gender feature, the two feature sets are the mirror images of each other in terms of default: default number and non-default case vs. non-default number and default case. (In {GDR:fem, NUM:pl, CASE: null}, case results as default by virtue of being unexpressed and nondistinct.) Defaultness will be explored more at length in section 5. Meanwhile, we see here a perfect illustration of the fact that homonymy is nearly always the easy solution that allows for no generalization and is to be retained only if and as long as nothing else works within reasonable limits.

The other possible account is to assume that the formal identity of singular oblique and plural in feminine nouns and adjectives is an instance of syncretism due to the effect of a rule of referral. Paraphrasing Stump's (2001: 46) informal statement, such a rule would say the following:

[10] *Null* means 'morphologically unexpressed'.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

(12) In the feminine noun class, singular oblique is inflected as plural is inflected.

This is different from a statement of homonymy. It is not the case that there are two identical exponents for two separate feature sets, but rather that, in order to know what the realization of a particular paradigm cell is, you are instructed to look up (referred to) another particular cell. In effect, there is one exponent, which leaves no space for homonymy.

Rules of referral work fine in cases like the Romanian conjugation (see Stump 2001: 213ff.). For instance, in the Present tense of 1st conjugation verbs, 3.SG and 3.PL are alike: cf. *invită* 's/he invites' or 'they invite'. Stump calls this type of syncretism 'nondirectional' because both forms involve the same affixation, namely of *-ă* (compare *invit* 'I invite'). Moreover, 3.SG and 3.PL constitute the natural class of 3rd person forms, so *-ă* may be analysed as an exponent of 3rd person not specified for number, which makes the syncretism 'unstipulated' because it has a rationale.

At first blush, the 'syncretism' of singular oblique and plural direct-oblique would seem to fall under the same category of identical affixation (of *-e/i*) to different forms. Yet, singular oblique and plural direct-oblique do not form a natural class. Their identical inflection would therefore have to be an instance of stipulated, i.e. semantically arbitrary syncretism, precisely like the identity of genitive-dative singular and nominative plural in Latin 1st declension. Not only would the Romanian feminine paradigm be descended from the Latin paradigm, but it would reflect it formally. But again this fails to address the crucial fact that singular oblique and plural direct-oblique do not constitute a natural class not simply because they are different, but because they are the **opposite** of each other in terms of the default or non-default values of the case and number features. There is thus a specific relationship, unlike in stipulated syncretisms, although not that of natural class as in unstipulated syncretisms.

Moreover, the theoretical significance of rules of referral derives from their involvement in complex inflectional systems consisting in several declension or conjugation classes, usually quite rich, and in which the same subset of cells turns out to be formally identical paradigm after paradigm. This is not at all the situation in Romanian, where only one paradigm is concerned, and it is a minimal paradigm contrasting two numbers and two cases, with the consequence that the observed 'syncretism' could not be other than what it is in quality or quantity without annihilating itself. To see this, consider paradigm (3) again:

(3)

	Singular	Plural
Direct	1 cas-ă	3 cas-e
Oblique	2 cas-e	4 cas-e

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

'Syncretism' involves cells 2, 3, and 4. Suppose it were qualitatively different, involving a different set of three cells. It could be 1/2/3 or 1/3/4. The first option branches according to whether *casă* or *case* is the spreading form. The former possibility (*casă* in 1, 2, and 3) would amount to (a) neutralizing the case contrast in the default singular while keeping it in the non-default plural, contrary to a widely attested tendency towards syncretising the non-default (marked) member, if at all; (b) neutralizing the number contrast in the default direct case while keeping it in the non-default oblique case, again in opposition with the same tendency. The latter possibility (*case* in 1, 2, and 3) would yield the same unfortunate results, and in addition it would mean inflecting the entirely default singular direct case.[11]

Consider now 1/3/4 syncretism with the same two possibilities as previously: *casă* in 1, 3, and 4 would neutralize the number contrast in the default direct case as well as the case contrast in the non-default plural number. *Case* in 1, 3, and 4 would spread the same havoc, and would even compound it by leaving singular non-default oblique (cell 3, now *casă*) uninflected, while inflecting all default singular direct.

Suppose now syncretism were quantitatively different, involving less or more cells. More cells means all four and the complete disappearance of all inflection, be it case or number. Less cells, on the other hand, means two alternate-world two-cells syncretisms: 1/2 and 1/3, since 2/3, 2/4, and 3/4 are effective in real life. Of these, 1/2 erases the case contrast in the default number, while 1/3 erases the number constrast in the default case, unnaturally in both instances.

Given all this, rules of referral in the case at hand seem both to miss a point (the mirror-image relation between the 'syncretised' members) and to be vacuous because what they intend to describe could not be different given defaultness relations the importance of which should by now be apparent. What we need is therefore an account of why it couldn't be different, and that is where toggle exponents come into play.

The toggle property and its consequences will be explored more in detail when we come to the actual analysis of Romanian declension. Let me just say for the present that, given this property, the *-e* ending of *fete* and the *-i* ending of *pisici* 'of/to (a) cat, cats' express feminine (i.e. non-default) gender always, and they express non-default number *or* case, not both. Although it involves defaultness crucially, toggle marking ought not to be counfounded with default marking. I am hinting at still another ill-thought account, this time along the lines of Halle's (2000) analysis of Latin declension in the Distributed Morphology framework.

Given the dearth of forms in the Romanian nominal inflection, such an account would have to assume two case-number exponents for feminine nouns such as *fată* : *-ă* defined as [+Direct, –Num], and *-e* [null], i.e. unspecified for any feature. The Subset Principle would then ensure that /-e/ is chosen whenever the

[11] Recall that final *-ă* is a word-class exponent, not an inflectional one.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

conditions for inserting its more specified competitor are not met. Obviously, this is doing little more than throwing theoretical garments over bare facts, namely that *fată* is the singular direct form, whereas *fete* takes on all other values. It does not even take in all the facts that must be brought to bear, e.g. that *-e* is not the only exponent of feminine plural (cf. *pisică* / *pisici* 'cat[s]'). Integrating this fact and more would then force us to add ever more declensional paradigms without increasing explanatory adequacy. Moreover, such a system results in the counterintuitive conclusion that plural is the default value.

Having thus outlined the analytical issues raised by the Romanian declension and refuted (I hope) accounts of them through homonymy, syncretism, or the Subset Principle, I now proceed to a presentation of the framework required for the account I will finally retain.

4. THE LEXICON-PHONOLOGY INTERFACE: THEORETICAL BACKGROUND

Although I adopt a rather standard view of the lexicon, the most explicit lay-out of which is to be found in HPSG theory (see Pollard, Sag 1994), the present framework is not bound to one general theory of language, but it is compatible, I believe, with all 'non-syntactocentric' theories that assign a central position to the lexicon (see, e.g., Sadock 1991; Jackendoff 1997, 2002; Ackema, Neeleman 2004, insofar as they build on Jackendoff's leading assumptions). It is irreconciliable, in contrast, with theories that consider word-formation a syntactic process, of which Distributed Morphology is now the dominant version (see Halle, Marantz 1993).

According to this view of the lexicon, words are represented as matrices consisting in features and their values. For instance, a concrete word-form such as English *cats* expresses a matrix that contains, among other things, a list of phonological features whose value is the sequence ⟨k, æ, t⟩, a categorial feature valued as *noun*, and a number feature whose value is *plural*.

It is a characteristic trait of word-forms, more conspicuous in certain language types, that they need not separately exhibit all the feature values the meaning of the form leads us to assign to them. Thus, *cats* exhibits its plurality through the dedicated exponent *-s*; the plurality of *mice*, in contrast, only appears in the global contrast with the singular form *mouse*. Historical facts apart, nothing would prevent the contrast from being reversed, with *mouse* the plural and *mice* the singular.[12] It does not take a thorough investigation, on the other hand, to ascertain that *cats* could not be the singular of which *cat* was the plural.

[12] True, the high vowel or diphthong always goes with the plural in such pairs (cf. *foot* / *feet*, etc.), which may establish a useful pattern for acquisition. This phonological contrast is only meaningful in such pairs, however, since obviously not all nouns with high vowels are plural (see *price*).

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

Of course, this observation, trivial in itself, makes sense only if the morphological theory that frames it belongs to the inferential-realizational family in Stump's (2001) typology, that is if we do not assume that *mice* contains a morpheme identical to *-s* but for the fact of being silent, the covert presence of which at the end of the word-form induces umlaut. Why inferential-realizational theories of morphology should be considered superior to lexical theories of the incremental or realizational sort, morphemic theories in other words, has been argued at length by several authors (see, e.g., Stump 2001).[13] Let me just make clear that 'morpheme' is not a term of the theoretical vocabulary I will be using. I use the term 'exponent' instead, meaning the **expression** by whatever device of the features present in the abstract sign of which the word-form is the actual realization.

Given this, the necessity of an interface between lexical matrices and phonological forms appears self-evident. More precisely, we need a level at which only those features of the matrix that perform some **morphological** function are present. I insist on 'morphological' because I do not want to imply that all such features are necessarily given a specific phonological realization. For instance, the nominal feature of *(a) walk* is not associated with any exponent in the word-form. It turns out to be morphologically functional, however, because it directly determines the paradigm the form belongs to, namely the two-cell paradigm *(a) walk* / *(many) walks*, rather than the four-cell verbal paradigm *walk* / *walks* / *walked* / *walking*. It must therefore be included at the interface level, although some formal means has to be devised in order to show that it is not phonologically realized. Moreover, the formalism should also be able to show that non-realization is not a matter of principle in this case, but it is a property of the particular grammar. In Bantu languages, where most nouns bear a noun class prefix, the nominal feature **is** associated with exponents (see Kihm 2005).

5. A FORMAL REPRESENTATION OF THE INTERFACE

As mentioned in the Introduction I call 'Concrete Lexical Representations' (CLRs) the formal objects that constitute the lexicon–phonology interface. As implied in the preceding discussion, the whole issue with CLRs can be summed up in two questions: (a) What feature values of a given lexical matrix are represented in the associated CLR? (b) How are these feature values formally arranged?

I submit that defaultness theory is the adequate tool to answer the first question. We saw in section 3 what crucial role defaultness considerations play in shaping Romanian paradigms. We must now be more specific. Features belong to two types. In one type, there is a default value that contrasts with one or several

[13] I wonder if 'lexical' is the right term to characterize these theories. All theories of morphology are lexical insofar as they need a lexicon.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

non-default values.[14] Number (NUM) is such a feature: singular is its default value. In two-number systems like English or Romanian, plural is non-default. In three-number systems like Ancient Greek or Arabic, there are two non-defaults: dual and plural.[15] In the other type, the diverse feature values do not stand in a defaultness relation. This is probably the case of the categorial feature (CAT), none of whose values – consisting at least of noun (N), verb (V), and particle (P) – seems to be default with respect to the other (on categories, see Blevins 2005).

A strongly supported empirical observation is that default feature values such as singular are not expressed by dedicated exponents. I therefore assume that **CLRs only include features with non-default values or not entering a defaultness relation**. In other words, they comprise precisely those feature values which are susceptible of phonological realization, thus representing 'compressed' signs with respect to the full lexical matrices.

A discussion cannot be eschewed at this point. Isn't there some circularity in defining as default and therefore unexpressed those feature values which mere observation shows us to be unexpressed? I don't think the suspicion is warranted. Take singular: it is default and unexpressed with respect to plural (and dual). We know of cases, however, where singular, then called 'singulative' is non-default and expressed with respect to an interpretation that may be defined as massive. An example is Classical Arabic *šajarun* meaning 'a collection of trees', next to singulative *šajaratun* 'a tree', from which the plural *šajaraatun* 'some trees' may be formed. The referential meaning of the word is crucial here: trees are entities that usually appear in clusters. This suggests there exist independent, to wit cognitive grounds to define a given feature value as default or not. Particular grammars take these grounds into account in diverse fashions and to varying degrees, this is an empirical issue. There is no circularity at any rate.

The second question: How are feature values arranged in CLRs? points to a crucial difference between lexical matrices and CLRs, namely that linear ordering (concatenation) has no status in the former, whereas it must be an essential property of the latter, since they constitute an interface the output of which is a phonological form. Bringing linear order in actually is a central function of CLRs.

Set theory seems to be a proper tool to achieve this function. That is to say, **I assume the elements of CLRs to be (ordered) lists and (unordered) sets**. Why the duality? Recall that one thing the formalism has to account for is why some feature values, although represented in the CLR because morphologically active, remain unrealized. Phonological realization requires ordering. Hence the conclusion that only those features values which are mutually ordered within

[14] This makes the difference between defaultness and markedness, as the latter is binary as a matter of principle (marked vs. unmarked), while the former need not be.

[15] Or there is a non-default non-singular consisting in default plural and non-default dual. Defaultness may be a recursive property.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

delineated domains, that is which belong to identified lists, can be associated with sequences (from 1 to *n*) of phonological segments.[16]

In every CLR describing a major category item (minimally a noun or a verb), there is a list that is identified by the attribute ROOT (ℜ) and has sequences of consonants and vowels as a value, which themselves receive specific segments as their values. CV sequences build up templates, in a more or less constrained fashion depending on the particular languages. All this is illustrated in (13) where the Latin root *mūr-* 'WALL' is represented :[17]

(13) ⟨ℜ C:m, V:u, C, V:u, C:r, V⟩ *mūr-*

Following Lowenstamm (1996), I take CV to be the basic unit of root templates. Some Cs and Vs do not receive a value. This is the case in (13) of the final V, the role of which will appear presently, and of the middle C under the assumption that surface long vowels should be broken up into a sequence of two identical vowels separated by a C slot with virtual features similar to those of the flanking vowels: /V:x C:x V:x/ → /Vxː/ (see Kihm 2003 for a motivated analysis of Classical Arabic long vowels along these lines).[18]

ROOT is subsumed by STEM (Σ), hence (14) for the stem *mūr-( )*, fully realized as *mūrus*, *mūrum*, *mūros*, etc.:[19]

(14) ⟨Σ ⟨ℜ mūr⟩ { }⟩ *mūr-( )*

I assume the stem to always consist in a list of length 2, at least in languages like Latin (see below). The empty set in (14) thus preempts the location of non-root or functional features possibly accruing to the root and filling up the stem. What features are these, and how are they organized?

In inflectional systems such as Latin one-to-one correspondence of features with exponents is the exception, as is well known, whereas cumulation is the norm. Let us take the word-form *mūros* as an example.The ending *-os* expresses the following feature-value pairs of the lexical matrix: NUM:pl, CASE:acc, GDR:masc, and DECL:II. In plain words, we are dealing with the plural accusative form of a masculine gender lexeme belonging to the second word-class or declension, to use conventional terminology.

[16] Notice that set-theoretic representations do not imply any hierarchy. I will return to this.

[17] I use Latin rather than Romanian as a language of exemplification because of its richer inflection. Roots' glosses are capitalized to suggest that, from a semantic viewpoint, roots may be equalled to lexemes.

[18] Interestingly, Classical Latin *mūrus* proceeds from an archaic form *moerus*, in which the C slot shows up as a glide.

[19] For simplicity's sake I will henceforth use conventional spelling for roots, unless there is need for a more precise notation.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

None of these features is expressed by a dedicated exponent. Rather, *-os* expresses all of them as a set. Notice, moreover, that masculine may be considered the default value for gender.[20] From the argument in section 2.1 it then follows that GDR:masc need not be included at all in the CLR. (See below for a more precise view of gender.) A similar reasoning applies to the plural value of NUM: if number was default singular, NUM could be left out entirely at the interface level. Since it is non-default plural, and the number contrast in binary in Latin, it is enough to include NUM in the CLR without specifying the value, as it can be no other than plural. Likewise, there are grounds, which I cannot detail here, for considering accusative rather than nominative the default case in Latin, meaning that CASE:acc should also be left out. DECL:II, in contrast, has to be included as it lies outside of defaultness relations.

I conclude that the CLR of *mūros* contains a two-member set {NUM DECL:II} compressing the four feature-value pairs of the lexical matrix. Defaultness, that is the automatic retrieval of a particular feature value simply because no other value is explicitly specified within a given value space, appears thus as one reason why not all feature-value pairs of the lexical matrix are found in the associated CLR.

What we still have to account for, however, is cumulation, namely the fact that both NUM and DECL:II are expressed by the lone exponent *-os*. Here, as already suggested, the assumption that they constitute a **set** is crucial: since they are not ordered with respect to each other, neither of them may be individually expressed; only the set itself may be, for it is ordered with respect to the root list since they both are members of the stem list. This is illustrated in (15):

(15) ⟨$_{\Sigma}$ ⟨$_{\mathfrak{R}}$ mūr⟩ { NUM DECL:II}⟩

Two informations are still lacking in (15): that *mūros* is a noun form; that it is a word. As already mentioned, the first information is carried by the N value of the category feature in the lexical matrix. Although no dedicated exponent expresses this value, it must be present in the CLR as it is neither default nor non-default.

How is it present? Two possibilities suggest themselves. According to one, N is part of the functional feature set, and *-os* expresses it along with NUM and DECL:II. The ending thus appears to be inherently nominal, and it transmits its feature to the form as a whole, in keeping with Williams's (1981) idea that suffixes are heads of their words.

[20] More accurately, feminine is the non-default for gender. Within default non-feminine, masculine is the absolute default, whereas neuter is the relative non-default. The absorption of the neuter by the masculine in the Romance languages supports this analysis. See Lowenstamm (to appear) for an analysis of gender systems in a conceptually similar, but formally different framework.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

For reasons I cannot detail, I will not adopt this solution, however, and I will assume that being a noun or a verb is a property of the word, not of one of its subparts. If follows that N in the CLR constitutes a singleton set not ordered with respect to the rest of the CLR, hence unexpressed. I show this in (16):

(16) $\{_{W} \langle_{\Sigma} \langle_{\mathfrak{R}}$ mūr$\rangle$ { NUM DECL:II}$\rangle$ {N}} *mūros*

What (16) describes is the word-form *mūros*, analysed as a set labelled W for word, whose members are the singleton set {N}, specifying the word as a noun, and the stem list, consisting in the root list followed by the functional feature set.

In accordance with the realizational approach to morphology adopted here, I assume that all lexical matrices are of type *word*. There are no lexical matrices for bound morphemes as in Distributed Morphology.[21] 'Word' and 'lexical matrix' are thus full synonyms. Lexical matrices include all the features semantically compatible with a given word type. For instance, CASE is an adequate feature for words of type *noun*, not for words of type *verb*. It follows that 'lexical matrix' is also synonymous with 'lexeme' if the latter is understood as the common denominator of all changeable feature values (see Aronoff 1994: 11).

Words are realized as one or several word-forms (see Matthews 1972, Chapter 2): for instance, *mūrum* and *mūros* are two word-forms of the word or lexeme that can be notated as MŪR, with partially different selections and valuings of the features that make up the word. As shown, such selections and valuings are represented at the interface level called CLR. CLRs are not words in the sense of lexemes, but word compressions preparatory to word-forms. Nor are they word-forms, since they have no material shape, and they include elements that will not show up in the word-form. I propose to define them as morphological words, since all the elements they do include perform some morphological function, to the difference of words-lexemes.

Notice that morphological words, *qua* set-theoretical objects, involve no internal hierarchy. In other words, representations such as (15) or (16) are **not** equivalent to trees or labelled brackettings (see Janda 1983 and Anderson 1992 for arguments against word-internal hierarchical structure). On the other hand, as just mentioned, it is not quite true that "An uncompounded word's morphological form is not distinct from its phonological form" (Stump 2001: 12). Both forms are very close to one another, but morphological words include more information than do word-forms. This is unsurprising since morphology deals with form (μορφή) *and* meaning, and there is not always enough form to express all the meaning.

Despite this, however, CLRs are easily converted into Stump's format for realization rules. Take (16) for instance. I take it to be equivalent to (17):

[21] Clitics are probably words, however.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

(17) RR $_{\text{N, DECL:II, NUM}}$ (<X, σ>) $=_{\text{def}}$ <X*os*, σ>, where X is *mūr-*

As formulated (17) differs from Stump's RRs only inasmuch as it does not include features or feature values that can be retrieved by default, namely GEN (gender) when masculine, CASE when accusative, and pl(ural) because including NUM is enough to show its value to be non-default. Realization rules thus formatted may therefore be seen as formalizing the operations through which lexical matrices are converted ('compressed') into interface representations. In turn, CLRs may be considered to represent the 'unfolding' of the righthand terms of the realizational equations, namely <X*os*, σ> in (17).

This is precisely where I diverge from Stump: in the present model, the outcomes of RRs are not word-forms, but morphological words. As the latter are written in such a way as to be straightforwardly realizable as the former, the difference is notational and often trivial. Yet I believe it is important because set-theoretic CLRs, to the difference of RRs and of course word-forms, provide an explicit formalization of cumulative exponence such as Latin -*os*.

Moreover, the lists and sets that form CLRs constitute morphological **sites** in the sense of Guerssel & Lowenstamm's (1990) analysis of Classical Arabic verbal derivation (also see Kihm 2006). CLRs thus provide an especially clear view of nonconcatenative morphological processes. And they do this, notice, in full agreement with the reasonable assumption that "There is no theoretically significant difference between concatenative and nonconcatenative inflection" (Stump 2001: 9).

We are now provided with the formal equipment to achieve an analysis of the Romanian declension that accounts for its properties as exposed in sections 3 and 4. This implies a thorough examination of the two categories for which Romanian nominals inflect, namely number and case. Yet, since the expression of these categories always cumulate with that of gender, the latter will be examined first. Two issues will be successively looked at: gender marking, which includes the question of the Romanian 'neuter', and gender agreement.

6. THE EXPRESSION OF GENDER

6.1. Gender marking

As argued at length in Kihm (2005), I assume gender to be one possible realization of the more general category Class, an inherent, distinctive property of nouns.[22] Nominals, in turn, consist in the combination within lexical matrices and

[22] I mean by this that only nouns are morphologically classified according to inherent properties of their *denotata*, such as sex, position in a natural taxonomy (humans vs. animals, etc.),

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

CLRs of non categorized roots with a feature N. Such an assumption – that roots of themselves have no category as nouns, verbs, etc. – is by far not an obvious one, and it probably has to be qualified in all sorts of ways. Yet it seems inescapable for inflected languages such as the Semitic and 'typical' Indo-European languages, which is why I adopt it here. Note moreover that my N crucially differ from Marantz's (1997) *n* ('little n') in being a **lexical** feature, not a syntactic element. Given this, Class can be equated with values N may assume, specifying the latter's nounness. Class is thus the set of all possible values of N.

As mentioned earlier, Romanian is traditionally provided with three genders: masculine, feminine and neuter, a highly inadequate division in any event, since it confuses two distinct properties: gender as a 'naturalist' classification based on sex (male vs. female vs. no sex) and gender as a grammatical classification yielding word classes (see Corbett 1991, 1998; Harris 1999). This is well-known, as it is well-known that discrepancies abound between the two properties as evidenced by such a Romanian noun phrase as *un paşă smintit* 'a crazy pasha' where agreement reveals *paşă* to be a masculine noun (compare ***o*** *fată smintit****ă*** 'a crazy girl') adequately referring to a male character, although the *-ă* ending and the definite article (*paş****a*** 'the pasha' like *fat****a*** 'the girl') put it in the feminine word class.

Some conceptual and terminological clarification is therefore in order. Considering that awareness of phenotypical sex (if any) pertains to world knowledge, I propose to call gender *qua* naturalist classification w-gender, whereas grammatical gender will be called g-gender. Such qualifications as masculine, feminine, and neuter ought then to be reserved for w-gender exclusively. For g-gender I will use the formal features F and non-F (see below).

As we saw in section 2, neuter cannot be considered a third gender in Romanian, neither as a word class – since so-called neuter nouns are in fact ambigeneric – nor as a natural class referring to, say, sexless or inanimate entities. I will therefore assume that nouns referring to inanimates are not valued for w-gender in Romanian, they only have a g-gender value. Nouns with animate reference, in contrast, bear both values, typically matching, so that feminine ≡ F, and masculine ≡ non-F, but sometimes not as we have just seen with *paşă* (and see below).[23] Thus, *colţ* 'corner' cannot be said to be masculine, it is merely non-F, whereas *prieten* 'friend' is masculine and non-F, and nonmatching *paşă* is masculine and F.

shape (flat vs. elongated objects, etc.), or other criteria still (for a review, see Aikhenvald 2000). Verbs are never classified in this way (e.g., dangerous vs. pleasant activities), as far as I can tell. Of course, there are verbs which, e.g., can only be used for animals, such as *whelp*. But this is a lexical and, to a large extent, cultural matter. No special portion of *whelp*, analogous to a gender exponent, points to the fact.

[23] As explained in fn. 8, Romanian singles out a syntactically identified 'human' w-subgender among animates. On the other hand, one may surmise that nouns referring to sexless or exotic animates such as *furnică* 'ant' or *elefant* 'elephant' are also unvalued for w-gender and only have g-gender (here F and non-F respectively).

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

G-gender thus appears as a morphological property related to the two possible values of N, N:F and N:non-F, defining two word classes, F and non-F. There are formal clues allowing one to allot each particular noun to one or the other class, as we shall see presently.[24] Ambigeneric nouns belong to both classes : they include N:non-F in the singular, and N:F in the plural, and they are most often without a w-gender value, although this cannot be taken as a defining characteristic. Now, given the logic of privative contrasts (cf. Trubetzkoy 1939/1969), non-F actually means no marking. In other terms, **non-F is the default value of N**, whereas F is its non-default value. I will therefore consider N to be unvalued rather than negatively valued in non-F nominals (see Farkas 1990: 543, fn. 9), so the contrast is between N:F and bare N. This is made more precise in the following.

Traditional grammars devote much space to the formal overt correlates of g-gender. Here I propose a generalization according to which singular nouns are divided into two groups : (A) nouns ending in a consonant or a root vowel ; (B) nouns ending in a desinential vowel. Group (A) includes the following forms with indication of g-gender ([±F] = ambigeneric) :

– **(A1)** nouns ending in a non palatalized consonant or legitimate consonant cluster (e.g., *plop* [non F] 'poplar', *tren* [±F] 'train', *urs* [non F] 'bear', *mosc* [non F] 'musk').

– **(A2)** nouns ending in root /I/ (i.e. [i] or [j]). Final root /I/ is most often unstressed, nonsyllabic ([ĭ]) or outright mute, and it palatalizes the preceding consonant to varying degrees : e.g., *ochi* [non F] [ok^{j}] 'eye', *arici* [a'ritʃ] [non F] 'hedgehog', *bici* [bitʃ] [±F] 'whip', *marți* [mart^{s}] [F] 'Tuesday', *puşti* [non F] [puʃt^{j}] 'brat'. It is syllabic, although unstressed, in a few borrowings (e.g. *cúli* [–F] 'coolie', *paciúli* 'patchouli' [F], *penálti* [±F] 'penalty kick').[25] It is also syllabic when (a) stressed (only in borrowings such as *pecarí* [non F] 'peccary') ; (b) the only vowel in the word, for which there is one example: *zi* [F] [zi] 'day'. It is realized as a glide [j] following a vowel (e.g., *pui* [non F] [puj] 'chicken', *pai* [±F] [paj] 'hay').

– **(A3)** nouns ending in a root vowel other than /I/.[26] Native terms must be distinguished from borrowings.[27] All vowels, stressed or unstressed, are attested in the latter: cf. *cóca* [non F] , *cafegibaşá* [non F] 'coffee vendor', *anşoá* [±F] 'anchovy', *baclavá* [F] *beizadé* [F] 'prince', *flamíngo* [non F], *cacáo* [F], *sombréro* [±F], *boleró* [±F], *acajú* [non F] 'mahogany', *atú* [±F] 'trump'. In native terms, in contrast, the only possible final root vowel other than /I/ is unstressed /U/,

[24] Adjectives in contrast receive gender through agreement.

[25] Whenever stress is relevant I show it by accenting the stressed vowel.

[26] The A2/A3 distinction rests on the special phonological properties of /I/. Otherwise, A2 and A3 are but two subgroups within the group 'ending in a root vowel'.

[27] The distinction is not entirely clear-cut. I am referring to those borrowings – from Turkish, French, English, etc. – the shape of which betrays they have not been fully integrated into the language from a morphophonological viewpoint (cf. Brâncuş 2004).

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

when the root would otherwise end in a *muta-cum-liquida* consonant cluster (e.g., *membru* [non F] 'member', *cioclu* [non F] 'gravedigger', *ministru* [non F] 'minister') or in a stressed root vowel (e.g. *léu* [non F] 'lion', *vizitíu* [non F] 'coachman'). That is to say, /U/ is an epenthetic vowel preventing native word-forms from ending in a Cr/l cluster or a stressed root vowel other than /I/. (This is of course a synchronic analysis ; historically, Romanian final /U/ continues the /U/ of the Latin declensions II and IV.[28]) Notice however that epenthetic /U/ shows up in the articulated form of a few nouns where it is not present, nor required, in the unarticulated form : cf. *cafeá* [F] 'coffee' vs. *cafeaua* [ka'fe̯awa] 'the coffee', *cazmá* [F] 'spade' vs. *cazmaua* 'the spade' (also see *zi* 'day' vs. *ziua* 'the day').

The data can be synthesized as follows : group A consists in singular nouns ending in one or more consonants, with or without /U/ insertion, or in a root vowel. They may be called singular non-desinential nouns, and they are overwhelmingly non-F.

N is therefore unvalued in these nouns (see above). Although unvalued, however, it must be present in the CLRs, for categorial features do not participate in defaultness relations (see section 5). No dedicated exponent expresses unvalued, bare N as the data make clear. It follows, according to the formal conventions developed in section 5, that bare N identifies a site unordered with respect to the stem within the word.

Actually, absence of value and absence of exponent (the latter formalized through unordering) must be seen as representing one and the same property. In other words, the claim that default N is never expressed as such seems to be a robust candidate to universality. Conversely, being valued and being expressed also count for one property.[29] This gives us the following CLRs for, e.g., *plop* 'poplar' and *culi* 'coolie', both consisting in a set labelled W whose members are the N set and a list Σ, whose members are the list ℜ made up of phonological segments (or the slots for such segments) and the empty set:

(18) $\{_{W}\ \langle_{\Sigma}\ \langle_{\Re}$ plop$\rangle\{\ \}\rangle\ \{N\}\}$
(19) $\{_{W}\ \langle_{\Sigma}\ \langle_{\Re}$ kúli$\rangle\{\ \}\rangle\ \{N\}\}$

What about F nouns in group A? We shall put aside the few items such as *paciuli* 'patchouli' and *cacao*, because their morphophonological integration turns out to be quite weak: for instance, *paciuli* never takes the singular article

[28] All masculine nouns of Latin origin now ending in a consonant ended in *-u* before the 16th century, as a reflex of Latin *-um* (see Bourciez 1910/1967: 555). Final *-u* was better preserved in Aromanian, the dialect spoken in Albania, Bulgaria, Macedonia, and Northern Greece.

[29] In noun class languages such as the Bantu languages, it may be that no value of N is default, so all nouns are somehow marked for a given class. In gender or word-class languages, in contrast, it seems always to be the case that one value of N is default and devoid of exponent. In Latin, for instance, that would be declension III (cf. *plebs* 'people').

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

(**paciulia*), but only the plural article (*paciulile* 'the patchoulis'), whereas *cacao*'s articulated form is *cacaua* 'the cacao', obviously through analogy with *cafea* 'the coffee' – see Lombard, Gâdei 1981, II: 30-33.) A more serious problem is raised, however, by items such as *cafea* 'coffee', *cazma* 'spade', *zi* 'day' and week-day names, e.g. *marți* 'Tuesday'. We shall return to them after we have examined group B nouns.

Group B is complementary to group A, as it comprises desinential singular nouns, i.e. nouns ending in a desinential, morphologically functional vowel *-ă* or *-e* such as *cămaşă* 'shirt' or *frate* 'brother'**.** All nouns ending in *-ă* are F. As we saw, there is a closed list of nonmatching items as far as the relation between g-gender and w-gender is concerned: e.g., *aghiuță* 'little devil' *papă* 'pope', *paşă* 'pasha', *popă* 'orthodox priest', *tată* 'father', *vlădică* 'bishop' – see Lombard, Gâdei 1981, II: 11-12). All other *-ă* nouns, insofar as they have w-gender at all – i.e. *fată* 'girl', but not *cămaşă* 'shirt' – are matching.

The same assumptions as were applied to group A entail that N:F combined with group B roots, being expressed by a dedicated exponent, namely the *-ă* or *-e* endings, identifies a site within the stem to the right of the root, corresponding to the empty set of (18) and (19). This I call the Righthand External Site (RES).[30] F nouns ending in *-e* (e.g., *carte* 'book') force us to posit two subvalues of F corresponding to two word classes within F nouns. Pursuing the logic of recursive defaultness, I will therefore assume that non-default F subsumes two values: a relatively default value notated Fa and realized as *-ă*, and a relatively non-default value notated Fe and realized as *-e*, Hence the following CLRs for *cămaşă* and *carte* :

(20) ⟨W ⟨Σ ⟨ЯR kəmáʃ⟩{N:Fa}⟩⟩
(21) ⟨W ⟨Σ ⟨ЯR kart⟩{N:Fe}⟩⟩

I will assume that items like *cafea*, *cazma*, *zi*, *marți*, etc. are assigned the same CLRs as in (20), hence, e.g., (22) and (23) for *cafea* and *zi* :

(22) ⟨W ⟨Σ ⟨ЯR kafɛ́a⟩{N:Fa}⟩⟩
(23) ⟨W ⟨Σ ⟨ЯR zi⟩{N:Fa}⟩⟩

The phonological component then sees to it that : (i) the *-ă* exponent of N:Fa is deleted after a stressed vowel : /kafɛ́a-ă/ → [kafɛ́a], /zí-ă/ → [zi] ; (ii) before other endings an epenthetic segment is inserted, whose form is /U/ before a low vowel (cf. *cafeaua* /kafɛ́a**U**-a/ 'the coffee', *ziua* /zí**U**-a/ 'the day'), but /l/ before a non-low vowel (cf. *cafelei* /kafé**l**-e-i/ 'of/to the coffee', *cafelele* /kafé**l**-e-le/ 'the coffees', *zilei* /zí**l**-e-i/ 'of/to the day', *zilele* /zí**l**-e-le/ 'the days').[31]

[30] 'External' because the framework also provides for root-internal sites, not represented in Romanian (see Guerssel, Lowenstamm 1990 ; Kihm 2003, 2006).

[31] Note that *zi* has a variant *ziuă*, which supports the present analysis (see Lombard, Gâdei 1981, II: 29, 145).

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

What about non-F nouns ending in *-e* such as *frate* 'brother', *pântece* 'belly', etc. The site N identifies cannot be unordered as with non-F nouns such as *plop*, etc., as it would make it impossible for an exponent to appear. We assumed that the Fa/e non-default value of N contrasts with no value, that is default, bare N. Now, by the same logic of defaultness that allowed us to bifurcate N:F into a relative default and a relative non-default, we can assume that default N also subsumes two values: unrealized absolute default N and a relative non-default which I label F– in order to point to the fact that nouns belonging to this word-class are non-F, but they are not unmarked for g-gender in the morphological sense of 'marked', since, as just proposed, being valued entails being expressed.[32] That the same exponent *-e* expresses the two relative non-defaults, one relative to the default value of N, the other to its non-default value, then falls out rather naturally. Hence (24) for *frate* 'brother':

(24) ⟨$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\Re}$ frat⟩{N:F–}⟩⟩

Henceforth I will use non-F to designate the union of bare N and N: F–, both contrasting with N:Fa/e.

Finally, I wish to insist that, since the neuter is not an independent category in Romanian, no account other than historical is required or feasible for the fact that some singular non-F nouns remain non-F in the plural while some others 'become' F. In other words, nothing allows us to predict that, given two F– nouns refering to inanimates, one, *burete* 'mushroom' (cf. *un burete bun* 'a good mushroom'), has the non-F plural *bureți* (cf. *bureți buni* 'good mushrooms'), but the other, *dulce* 'sweet' (cf. *un dulce bun* 'a good sweet') has the F plural *dulciuri* (cf. *dulciuri bune* 'good sweets'). As we shall see, such a discrepancy tells a lot about the nature of the pluralization process.

6.2. Gender agreement

A noun's w-gender (if assigned) – e.g., that *paşă* 'pasha' refers to a male human being – is only manifested through the agreement phenomena related to that noun (see Corbett 1998). This is what makes a study of agreement necessary for a full view of Romanian gender, especially considering the possible mismatches between w-gender and morphologically marked g-gender (see above). Here I will use a conceptually and technically minimal theory of agreement, namely that it consists in feature value-sharing between a controller (the head of the noun phrase for our concerns) and one or several controllees.[33]

[32] The presence of nouns like *pântece* 'belly' in this word-class shows we are not dealing with a mismatch case as with *paşă*. Moreover, contrary to *paşă*, *frate* is articulated as a non-F noun (cf. *fratele* 'the brother').

[33] See the following section for complex facts of agreement with ambigeneric nouns.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

Two controllees are considered here : the (definite) article and the indefinite singular quantifier *un*/*o* ‘a (masculine/feminine)’. The agreement properties of the latter are the same as those of other quantifying elements that may be inserted into the left-side domain of the noun phrase (e.g., *alt* ‘other’, *câtva* ‘some’, etc.) and for attributive adjectives (generally postposed), which exempts us from looking at these items.

Given the enclitic character of the Romanian article, I analyse it as the realization of a feature D (‘determiner’) identifying a site to the right of the stem within an extension W’ of the word. W’ can be viewed as formalizing the notion ‘phonological word-form’ when it is distinct from the word-form proper, thus allowing for clitic insertion. The enclitic exponent of the article is attached to the head of the noun phrase or to a preposed adjective: cf. *prietenul sărac* ‘the poor friend’ vs. *săracul prieten* ‘id.’ (see Cornilescu 1992 for a syntactic account ; also see Pană Dindelegan 2003).[34] Abstracting from number and case, the following CLRs are therefore assigned to *prietenul* ‘the friend’, *fratele* ‘the brother’, *fata* ‘the girl’, and *cartea* ‘the book’:[35]

(25) ⟨$_{W'}$ {$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\text{Я}}$ prieten⟩{ }⟩ {N}}{D}⟩ *prietenul* /prieten=ul/
(26) ⟨$_{W'}$ ⟨$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\text{Я}}$ frat⟩{N:F–}⟩⟩{D}⟩ *fratele* /frat-e=le/
(27) ⟨$_{W'}$ ⟨$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\text{Я}}$ fat⟩,{N:Fa}⟩⟩{D:Fa}⟩ *fata* /fat=a/
(28) ⟨$_{W'}$ ⟨$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\text{Я}}$ kart⟩{N:Fe}⟩⟩{D:Fa}⟩ *cartea* /cart-e=a/

In (25), bare N is unordered in the word set and D agrees with it in the sense that it is given no specific gender value, it is gender-default in other terms. Gender-default singular D is realized as *-l* ; the /u/ that precedes is the same epenthetic vowel as we saw in *membru*. In (26), the RES is filled by N:F–. D, still valueless, is realized as *-le*, a variant of *-l* given the following phonological representations in accordance with Lowenstamm’s (1996) CV format :

(29) [pri'jeten.l.]
(30) ['fratel.]

The dots mark the vowel slots that must be identified from left to right. In (29) the slot following [n] is identified by epenthetic [u], while the slot following [l] is allowed to remain empty. In (30) the only underlyingly empty slot follows [l], and it is identified by copying final *-e*.

[34] The possible syntactic causes of the article’s cliticization are irrelevant at the level of the present analysis. There is the same semantic difference between *prietenul sărac* and *săracul prieten* as there is in French between *l’ami pauvre* and *le pauvre ami*.

[35] As we shall see later on, (25)-(28) are well-formed CLRs for singular, direct case nominals.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

A significant upshot of the analysis of -*le* as an allophone of -*l* triggered by final non-F -*e* is that D actually agrees with the **absolute** default value of N. We shall see presently that F nominals confirm this conclusion. Also notice that (29) and (30) describe (extended) word-forms, implying that determination (i.e. article cliticization) in Romanian is primarily a morphological matter, perhaps something syntax hasn't to take care of at all.

Finally, in (27) and (28), D is associated with the same exponent -*a* in both cases, which means it receives default Fa as a value, in support of the analysis of the -*l*/-*le* alternation above. Phonological processing then reduces the virtual [ăa] sequence of (27) to [a]. The /ea/ sequence of (28) is not affected, except that [e] loses part of its sonority, becoming the open glide [e̯].

Recall now that nouns refering to inanimates are not assigned w-gender, only g-gender. Nouns refering to animates have both, and a default implicature from g- to w-gender may be posited such that F nouns refer to females, while non-F nouns refer to males. The implicature does not hold in two cases. First, there are epicene nouns such as F *pisică* which, in addition to denoting the whole species, may refer to she-cats as well as to tomcats. No particular grammatical consequences follow from this semantic phenomenon whereby the female is taken as representative for the species – as in German *(die) Katze* or Bulgarian *kotka(ta)*, and contrary to French *(le) chat*.

The second exception is grammatically more significant. I am referring to the already mentioned nonmatching nouns which are F for g-gender, but masculine for w-gender, such as *paşă* 'pasha', *vlădică* 'bishop', etc. As far as agreement is concerned, these nouns show a special property. Whereas g-gender controls agreement of the article in the singular (cf. *paşa* 'le pacha', not **paşul*), w-gender controls agreement of the indefinite quantifier and the adjective : cf. *un paşă smintit* 'a crazy pasha' (**o paşă* **smintită*), *paşa smintit* 'the crazy pasha' (*paşa* **smintită*). It also controls agreement of the plural article : cf. *paşii* /paʃ-$i_{PL=}i_{D,\ non\text{-}F,PL}$/ 'the pashas', not **paşile* */paʃ-$i_{PL=}le_{D,\ F,PL}$/ (see below for /i_{PL}/). A few nouns ending in -*e* belong to the nonmatching group as well, such as *bade* 'father [religious]', *haple* 'nincompoop', *gâde* 'hangman'. One thus finds *gâdea* 'the hangman', but *un gâde smintit* 'a crazy hangman', and *gâzii* 'the hangmen' (see Lombard & Gâdei 1981: II 7-8, 54-60). Given the present analysis, they unambiguously enter the Fe word class of *carte(a)* '(the) book', with their nonmatching character their only peculiarity.[36]

An interesting comparison at this point, which must be put off for future research, is with French, where nonmatches are even rarer than in Romanian, and g-gender is the sole controller of agreement: cf. *la sentinelle folle* 'the crazy sentry', not *la sentinelle *fou* (although up to a recent period all sentries were males).

[36] These nouns are frequently regularized into the F– word class, hence *gâdele* 'the hangman' like *fratele* 'the brother' (see GA 1963 : 85).

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

The discrepancy between article agreement and other agreements in Romanian is clearly related to the strongly local character of the noun-article relationship, inasmuch as the article shows up as an enclitic that must be integrated as soon as the CLR, i.e. in morphology. The other modifiers, in contrast, are syntactically combined (*merged* in minimalist terminology) with the noun. I cannot proceed any further on this issue here.[37] It would seem necessary, however, at least to draw a terminological distinction between both types of relation. Following Wechsler & Zlatić (2003), I call **concord** the local relation involving g-gender, and I call **agreement** the not so local relation involving w-gender.[38] We shall therefore say that the article concords with the noun for the g-gender grammatical feature (N's value), whereas the other modifiers agree with it for the w-gender encyclopaedic property through the implicature masculine ⊃ non-F. Nonmatchers, F but referring to masculine entities, violate the implicature. (Note there do not exist, to the best of my knowledge, opposite nonmatchers, that is non-F nouns referring to feminine animates.[39] If this is a fact, it means that only the non-default g-gender may violate the implicature.)

The concord vs. agreement contrast is only visible in nouns referring to animates, therefore provided with w-gender, where the value of the latter happens not to match with the g-gender value for the same nouns. Although they are admittedly few, they turn out to be crucial for the analysis. With nouns denoting inanimate entities, agreeing items have no w-gender specifications to look for. Agreement and concord are then non-distinct, referring back to the same feature set.

A final hitch that must be mentioned is discussed by Farkas (1990): when a demonstrative pronoun refers back not to an NP that could assign it a gender through agreement, but to an event that cannot, it appears under its F form, but adjectives predicated of it are non-F. The relevant example is the following (Farkas's [9], p. 541):

(31) Petru e acasă. Asta e uluitor / *uluitoare.
Peter is at.home this$_{F}$ is amazing$_{non\text{-}F}$ / *amazing$_{F}$
Peter is at home. This is amazing.

Farkas's account for this puzzling phenomenon is that the demonstrative pronoun *asta* gets F g-gender through a special default rule for antecedentless pronouns that overrides the more general rule according to which non-F is the default. This special value is not carried over to the predicative adjective, though,

[37] See below for the difference between the singular and the plural article.

[38] Insofar as the article is part of the word, the concord vs. agreement contrast may be viewed as related to the distinction between paradigmatic and syntagmatic features (see Farkas 1990).

[39] Epicene nouns such as non-F *medic* 'doctor', which may refer to women as well as to men, or ambigeneric *model* in the sense of 'fashion model' of course do not count (see *pisică* above).

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

which therefore appears with the general default g-gender value, that is non-F. Farkas then goes on to suggest that the reason for non-transmission may be that the special F value is a paradigmatic feature that lacks a syntagmatic counterpart.

In our terms (see fn. 38) this means that the pronoun's F value results from concord, whereas the adjective's non-F is the result of agreement for a default value. We might thus be tempted to conclude that event-denoting *asta* actually stands for (is the exponent of) a feature set that includes unvalued N (as it must, since events are kinds of entities). Agreement proceeds normally, but for some reason I will not speculate about unvalued N is locally interpreted as F.

7. THE EXPRESSION OF NUMBER

Since Romanian does with two numbers, singular and plural, I will view plurality as the interpretation of a functional feature Number (NUM) possibly present in a nominal's lexical matrix and CLR. This means treating pluralization as inherent inflection (see Booij 1996) or even derivation (see Beard 1995 and below), thus removing the need for a Number projection (NumP) in syntax. If NUM enters the CLR, the resulting nominal is interpreted as plural; if it does not, the interpretation is singular by default.[40] We thus achieve a formal expression of the fact that the singular is morphologically unmarked and of the assumption that it is semantically default vis-à-vis the plural.

Number is inherently linked to gender in Romanian in the sense that all the exponents that realize the former also realize the latter: for instance, final *-i* ([ĭ]) in *prieteni* '(some) friends' and *plopi* '(some) poplars' seems to contrast as a non-F plural marker (respectively masculine and w-genderless) with the F plural markers *-e* of *fete* '(some) girls' and *-uri* of *gheață* / *ghețuri* 'icicle(s)', respectively feminine and w-genderless.

Yet, such a reading off of properties from exponents is never a matter of course in a language like Romanian that shows so little inclination for one-to-one correspondence of form and meaning. Indeed, *-i* cannot simply be the plural ending of non-F nouns since it appears on many Fa/e nouns referring to animates or to inanimates : cf. *inimă* / *inimi* 'heart(s)', *pisică* / *pisici* 'cat(s)', *gară* / *gări* 'station(s)', *vulpe* / *vulpi* 'fox(es)', *carte* / *cărți* 'book(s)', etc. Moreover, as we know, many non-F nouns in the singular – the conventional 'neuters' – take the F endings *-e* or *-uri* in the plural : cf. *ac* / *ace* 'needle(s)', *cuvânt* / *cuvinte* 'word(s)', *dulce* / *dulciuri* 'sweet(s)', *fir* / *fire* 'thread(s)', *val* / *valuri* 'wave(s)', *vreme* / *vremuri* 'time(s)', etc.

On the other hand, no obvious nonmatch of g- and w-gender is observed in the plural since, as we saw, the unarticulated plural of *paşă* is *paşi* '(some) pashas'

[40] Naturally, this tells us nothing about the actual interpretation of the singular and the plural (oneness, genericity, etc.), a distinct matter which I do not address here.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

similar to *prieteni* ‘(some) friends’, not **paşe* after *fete* ‘(some) girls’ ; likewise for *gâde* / *gâzi* ‘hangman/men’. True, *-i* also goes with F nouns (cf. *pisică* / *pisici* ‘cat(s)’), but articulation – *pisicile* ‘the cats’ with the F plural article *-le* – reveals *pisică* for a ‘real’ F noun in the plural as in the singular, contrary to *paşă* and *gâde* – cf. *paşii* ‘the pashas’, *gâzii* ‘the hangmen’, with the non-F plural article *-i*. We are thus led to suspect that nouns like *paşă* represent one more case, symmetrical to ‘neuter’ nouns, of g-gender ‘alternation’ according to number : nonmatchers in the singular because they are F, they are no longer so in the plural because they are then non-F, hence the non-F plural article *-i*, in keeping with the concord relation that holds between the noun and the article (see above). Our task now is to formalize this notion of ‘g-gender alternation’, the domain of which appears to be quite large in Romanian.

To this end we should first discard the conventional notion that a noun’s plural is formed from the corresponding singular. Rather we shall consider the singular and the plural as being two parallel formations from the same root or variants of the same root. And we hold it to be a contingent fact that in many languages the root is *de facto* indistinguishable from the word-form that gets interpreted as singular, thus making it look as if the plural was derived from it (cf. Turkish *baş* / *başlar* ‘head(s)’ – or English *head(s)* for that matter). This view is in perfect keeping with the already mentioned hypothesis that noun pluralization is a derivational or at least inherent-inflectional process.

Moreover, there is no lack of evidence in Romanian that this is the correct view. Several cases can be observed, for instance, where the singular and the plural are formed from two parallel versions of the root. Thus, *păturică* [pətu'rikə] ‘small blanket’, the diminutive of *pătură* ['pəturə] / *pături* ['pəturĭ] ‘blanket(s)’, has *păturele* for a plural, formed not from the derived root *păturic-*, but from another derived root *păturel-* (cf. *păturelele* /păturél-e-le/ ‘the small blankets’). Similarly one finds *lopățică* / *lopățele* ‘trowel(s)’ (cf. *lopată* / *lopeți* ‘spade(s)’), *tufănică* / *tufănele* ‘chrysanthemium(s)’. Consider also such examples as *zi* / *zile* ‘day(s)’, *baclava* / *baclavale* ‘baklava(s)’, *canava* / *canavale* ‘canvas(ses)’ (French *canevas*), *nuga* / *nugale* ‘nougat(s)’, *za* / *zale* ‘stitch(es)’, the plurals of which are formed from a root ending in /l/ which appears nowhere else (cf. *ziua* ['ziwa] ‘the day’, *baclavaua* [bakla'vawa] ‘the baklava’, etc., with epenthetic /U/).

Nor is it rare for more than one plural to be associated with one singular. It may be a matter of stylistic variation as in *aripă* / *aripi* ~ *aripe* ‘wing(s)’, *stradă* / *străzi* ~ *strade* ‘road(s)’, etc., the first variant deemed *recomandabilă* by prescriptive grammarians (see GA : 66-67); or each variant may be specialised in meaning: e.g., *roată* ‘wheel’ has plural *roți*, except in the idiom *a pune bețe în roate* ‘to put a spoke (lit. ‘sticks’) in the wheel (lit. ‘wheels’)’. Likewise *minut* has plural *minute* in the sense of ‘minute(s)’, but *minuturi* in the sense of ‘fast-food restaurant(s)’. In still other cases the plural shows some additional meaning lacking in the singular : e.g., *apă* only means ‘water’, but *ape* means ‘(kinds of) waters’ or

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

'reflections, glint'; *lapte* means 'milk', whereas *lapți* means '(kinds of) milks' or 'roe'. Such facts are reminiscent of Arabic broken plurals, for which the parallelism hypothesis seems inescapable (see Ratcliffe 1998 ; Kihm 2003, 2006 ; also see Halle 1973 about similar facts in English and Russian).

In the present framework, these facts suggest the following empirical generalization :

(32) The root's form and/or N's value may vary according to whether NUM combines or not with the root-N combination.

In simple cases, there is no difference. The pair *prieten* / *prieteni* 'friend(s)' may thus be represented as follows :

(33) $\{_{W}\ \langle_{\Sigma}\ \langle_{\Re}$ prieten$\rangle\{\ \}\rangle\{N\}\}$
(34) $\{_{W}\ \langle_{\Sigma}\ \langle_{\Re}$ prieten$\rangle\{NUM\}\rangle\{N\}\}$

NUM 'by itself', that is not sharing a set with another feature, is realized as *-ĭ*. In other words, *-ĭ* expresses 'plain' plural.

In *prietenii* 'the friends', D combines with NUM as shown in (35) :

(35) $\langle_{W'}\{_{W}\ \langle_{\Sigma}\ \langle_{\Re}$ prieten$\rangle\{NUM\}\rangle\ \{N\}\}\{D\ NUM\}\rangle$

NUM thus appears twice: by itself to the immediate right of the root, being realized as *-ĭ*, and then, through concord, in the D site/set. Since D and NUM are not ordered with respect to each other they show up as one cumulative exponent, whose form is *-i*, the plural non-F article. (And recall that *-ĭi* is pronounced [i].)

Another simple case is that of F nouns ending in *-ă* or (rarely) *-e* in the singular and in *-e* in the plural (e.g., *casă* / *case* 'house(s)', *iesle* / *iesle* 'crib(s)').[41] If *-ĭ* expresses plain plural, then *-e* has to express plural plus something, namely Fa/e, the non-default g-gender : *-e* = {N:Fa/e NUM} identifying the RES as shown in (36) :

(36) $\langle_{W}\ \langle_{\Sigma}\ \langle_{\Re}$ kas$\rangle\{N:Fa\ NUM\}\rangle\rangle$ *case* 'houses'

As for the articulated plural *casele* 'the houses' (id. *ieslele* 'the cribs'), it is assigned the same CLR as *prietenii*, modulo the differences in g-gender and concord:

(37) $\langle_{W'}\langle_{W}\ \langle_{\Sigma}\ \langle_{\Re}$ kas$\rangle\{N:F\ NUM\}\rangle\rangle\{D:F\ NUM\}\rangle$

[41] Nouns like *iesle* are only apparently invariable (see Lombard, Gâdei, II: 24, 103).

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

{D:F NUM} is realized as *-le*. Nouns like *zi* / *zile* ‘day(s)’ ou *cafea* / *cafele* ‘coffee(s)’ are analysed in the same way (see [22]-[23]). Note the F plural article is always *-le*, so there is no need to specify a value for F when it is a feature of D.

Let us turn to more complex cases. First we have nouns such as *pisică* ‘cat’ or *carte* ‘book’, which form the plural with plain *-ĭ* (*pisici*, *cărți*) like non-F *prieten*, but take the F plural article *-le* (*pisicile* ‘the cats’, *cărțile* ‘the books’) in accordance with their F g-gender. The problem raised by such nouns is how to account for their showing a plain plural although N is valued (as Fa/e) in their CLRs. This is where the generalization formulated in (32) finds its first opportunity to apply. Take singular *pisică*’s CLR :

(38) ⟨$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\Re}$ pisik⟩{N:Fa}⟩⟩

Carrying this pattern over to the plural would make *pisici* underivable since NUM would have to share the RES with N:Fa and to be realized as *-e* (see [36]). In order for the plain plural *-ĭ* to show up, N must be unordered, an option that is precluded for valued N:Fa. I therefore propose that in accordance with generalization (32), the CLRs of *pisică* and *pisici* diverge in the following way: N, valued Fa in the singular, takes on the less specified value F in the plural. Just like bare N, N:F has no exponent associated with it (only N:Fa/e does) so it is unordered, with the consequence that NUM does not combine with it. *Pisici*’s CLR is therefore similar to *prieteni*’s CLR :

(39) {$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\Re}$ pisik⟩{NUM}⟩{N:F}}

The articulated plural *pisicile* ‘the cats’ is easily accounted since, as we saw, the underspecified value F is the one D combines with (see [37]).

The same assumptions will account for Fe nouns such as *carte* / *cărți* ‘book(s) and non-F nouns such as *frate* / *frați* ‘brother(s)’. In the latter N is valued F– in the singular and it is realized as *-e*; when NUM enters the CLR, N:F– reduces to N and becomes unordered. The articulated plural *frații* ‘the brothers’ is non-distinct from *prietenii* ‘the friends’, and likewise nonmatchers like *paşă* and *gâde*, except that N’s value is Fa in the singular controlling the *-a* realization of the article. Singular N:Fa contrasts with bare N in the plural, hence *paşi(i)* ‘(the) pashas’.

Finally, ambigeneric nouns illustrate the opposite move. Take, e.g., *băț* / *bețe* ‘stick(s)’. The singular CLR is as in (40) :

(40) {$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\Re}$ bət^{s}⟩{ }⟩{N}}

That is to say, *băț* is non-F, non-distinct from *prieten*. In the plural, i.e. with NUM added to (40), NOM takes on the value Fa, hence (41) for *bețe* :[42]

[42] Whether actually Fa or Fe is undecidable, but it must be one or the other since, as we just saw, plain F would entail unordering.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

(41) ⟨$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\Re}$ bət^{s}⟩{N:Fa NUM}⟩⟩

The articulated plural *beţele* ‘the sticks’ follows.

Let me say it once again: nothing about non-F *băţ* allows us to predict that it has F *beţe* for a plural rather than non-F **beţi*. That is something that has to be learned.[43] Learning it, on the other hand, clearly involves somehow adding the information to the lexeme’s internalized representation. Evidence for this comes from agreement facts such as the following:[44]

(42) Am cumpărat două scaune. Pe un-ul dintre ele l-
I.have bought two$_{F}$ chairs PE one$_{non\text{-}F}$-the from them$_{F}$ it$_{nonF}$
am pus în sufragerie
I.have put in living-room
I bought two chairs. One of them I put in the living-room.

Ambigeneric *scaun* ‘chair’ is F in the plural and the numeral meaning ‘two’ takes on the F form *două* accordingly (compare *doi băieţi* ‘two boys’). In the connected sentence, however, both the nominalized numeral *unul* ‘the one’ and the proclitic object pronoun *l-* ‘it’ must be non-F because they refer back to the (implicit) singular which is non-F (cf. *scaunul* ‘the chair’, *un scaun* ‘a chair’). *Ele* ‘them’ in contrast is F as it refers back to plural *scaune*. Therefore, although only one of the two N values of an ambigeneric noun is ever realized at a time, both are accessible, which implies that ambigenericity is a property of the overall lexeme of which the singular and the plural are forms.[45]

The ‘mixed’ character of Romanian ambigeneric nouns is further supported by the fact that predicative adjectives qualifying mixed-gender conjuncts appear in the plural F form, as in the following example from Lombard (1974: 98):

(43) Perete-le şi poart-a sunt văruite.
wall$_{non\text{-}F.SG}$-the$_{non\text{-}F.SG}$ and door$_{F.SG}$-the$_{F.SG}$ are whitewashed$_{F.PL}$
The wall and the door are whitewashed.

Peretele ‘the wall’ is non-F singular, *poarta* ‘the door’ is F singular, *văruite* ‘whitewashed’ is F plural as if it agreed with one plural ambigeneric noun (cf. *Restaurantul e văruit* ‘The restaurant is whitewashed’ vs. *Restaurantele sunt văruite* ‘The restaurants are whitewashed’).

[43] Again, ambigeneric nouns like *nume* / *nume* ‘name(s), noun(s)’ are only apparently invariable. In the articulated singular *numele* ‘the name’, *-le* is the variant of the non-F article otherwise realized as *-(u)l* (see above); in the articulated plural *numele* ‘the names’, homophonous *-le* is the plural F article. Agreement resolves such ambiguities.

[44] I am grateful to Paolo Acquaviva (p.c., 03/30/05) for bringing these crucial data to my attention.

[45] Romanian ambigeneric nouns thus contrast with Italian feminine plurals such as *le mura* ‘the walls’ paired with *il muro* (see last section).

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

To summarize, Number or plural has two realizations in Romanian : as plain NUM (including underspecified {N:F NUM}) with *-ĭ* for an exponent, and as {N:Fa/e NUM}, realized as *-e* or *-uri*.

Before we turn to case I wish, for the sake of completeness, briefly to examine a few ambigeneric nouns which present peculiarities in terms of number marking (see Lombard & Gâdei 1981: II 37ff.).

First, there are a very few nouns like *buzunar* 'pocket' which may form an apparently non-F plural *buzunari* next to the more usual F plural *buzunare*, but which nevertheless count as ambigeneric because the *-ĭ* plural controls F concord of the article (*buzunarile* 'the pockets'). They do not endanger our analysis, however, since we know that *-ĭ* is a possible ending for F nouns (cf. *pisici* 'cats'). What we have to do, then, is assign *buzunari* the following CLR similar to (39):

(44) $\{_{W}\langle_{\Sigma}\langle_{\Re}$ buzunar$\rangle\{$NUM$\}\rangle$ {N:F}$\}$

The difference between the singular and the plural CLRs thus involves changing N's value from nil to underspecified F which does not enforce ordering. It might be that such a change, although apparently slighter than the one involved in *pisică* / *pisici*, is actually more serious, for it gives an ambiguous outlook to the gender alternation it entails – since *buzunari*, agreement and concord aside, could be the non-F plural of non-F *buzunar*. This would explain the rarity of such forms. Non-F nouns ending in a root vowel or glide like *studiu* ['studiu] 'study' (*studiul* 'the study') and forming an F *-ĭ* plural (*studii* ['studi] 'studies', *studiile* ['studile] 'the studies') are probably amenable to the same account.

Equally rare, actually limited to this one item, is the case of *zece* 'ten', where *-e* realizes N:F– in the non-F singular (*zecele* 'the ten'), and N:F– 'becomes' N:F in the plural (*zeci* 'tens', *zecile* 'the tens').[46]

8. THE EXPRESSION OF CASE

As already mentioned (see section 2), Romanian is the only Romance language that kept something of the Latin case inflection or declension outside the pronouns. Yet, as we saw, only noun modifiers such as the (definite) article and various adjectival quantifiers such as *un* 'a(n)', *alt* 'other', *fiecare* 'each', *cutare* 'such', *tot* 'every', *un* 'a', *vreun* 'some', and pronominals such as *altul* 'another', *totul* 'everything', *unul* 'a certain one', *vreunul* 'someone', always inflect according to a paradigm that contrasts two cases : default, morphologically

[46] *Doi* '2 (non-F, contrasting with *două* '2_{F})', *trei* '3', *cinci* '5', *opt* '8', and *zero* '0' are also ambigeneric and they form their plural with *-urĭ* (e.g., *treiurile* 'the threes'). The remainder *unu* '1', *patru* '4', *şase* '6', *şapte* '7', and *nouă* '9' are supposed to be ambigeneric as well, but they do not pluralize, so one cannot really tell.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

unrealized nominative-accusative or direct case vs. non-default, realized genitive-dative or oblique case. The following tables shows the declensions of *alt* 'other' and *altul* 'another' for both genders and numbers:

(45)

	Singular		Plural	
	non-F	F	non-F	F
Direct	alt	altă	alț-i	alt-e
Oblique	alt-ui	alt-ei	alt-or	alt-or

(46)

	Singular		Plural	
	Masc.	Fem.	Masc.	Fem.
Direct	altul	alta	alții	altele
Oblique	altuia	alteia	altora	altora

Given a feature system like that proposed by Halle (2000: 133-134) for Latin, the merger of genitive and dative into a unique oblique case is in the order of things. Halle assumes three Boolean features, oblique, structural, and superior, such that [–oblique] items are arguments of the verb (subjects or direct objects), [+oblique] items are non-arguments (indirect objects or adjuncts); [+structural] items are assigned case because of syntactic position, [–structural] items on semantic grounds; [–superior] items stand in governed positions, [+superior] items do not.[47] Genitive and dative are thus described by the feature sets [+oblique, +structural, –superior] and [+oblique, +structural, +superior]. The ungoverned character of dative, whether it can be defended for Latin or not, is certainly inapplicable to Romanian where indirect objects are clearly governed by their verbs. Therefore, dative is [–superior] as well and nondistinct from genitive. As for the nominative and the accusative, which do differ in terms of government, but are both [–oblique] and [+structural], it is a well-known fact that they fell together very early due to phonological deletion of their characteristic endings /-s/ and /-m/, on the one hand, and the loss of free word order in favour of SVO, on the other hand.[48]

Paradigms (45) and (46) ought to be compared with the paradigms of nouns given in (3) and (4), which I repeat here for convenience:

(47)

	Singular	Plural
Direct	cas-ă	cas-e
Oblique	cas-e	cas-e

[47] Halle is cautious to call his system 'provisional'. In the present framework, to the extent they are valid, these features ought to be seen as lexical features, the combination of which receives the case label. Working out such a conception lies well beyond the limits of this article.

[48] Earlier in the Eastern Romance languages, i.e. Italian and Romanian, than in the Western languages where -s proved more resilient.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

(48)

Singular	Plural
prieten	prieten-i

Paradigms (45)-(46) appear as well-behaved inflectional paradigms, almost canonical in the sense of Corbett (2005): the entirely default singular non-F direct cell contains a form that is nondistinct from the root-stem (plus something that looks like the definite article in [46]); all other cells exhibit distinctive exponents. Paradigm (48) is also canonical: it is the minimal paradigm for nouns that contrast in only one feature, namely Number. Only paradigm (47) is odd, as already pointed out: the singular oblique cell in it is identical to the two plural cells – if there are two cells in the plural.

This fact and the comparison of (45)-(46) with (47)-(48) raises the following four questions:

a. Why do nouns never inflect fully for case to the difference of modifiers and pronominals?
b. Do non-F nouns, which do not inflect overtly for case at all, include unexpressed case or do they exclude case altogether?
c. Why do only F nouns overtly inflect for case?
d. Why do they inflect as they do?

Answering these questions amounts to a full (if not final) account of Romanian declension.

8.1. Why do only modifiers and pronominals fully inflect for case?

The problem we are facing is to take stock of the difference between nouns and modifiers / pronominals, while registering what they have in common. Feature hierarchies seem to be the right tool to achieve this.

In the foregoing developments I assumed a feature N that all items of type *noun* share (also see Sag, Wasow 1999). Now I will assume that items such as *alt* also include modifier and pronominal features which I won't try to specify further, but will write down under the symbol MP (for modifier and pronominal). Items such as *prieten* and *fată* do not share these features, they only have N. For convenient reference, I will call the former items 'nominals', and the latter items I will call 'nouns'. The union of nominals and nouns (and standard adjectives) constitutes the category of substantives (see Blevins 2005).

The one difference that separates nominals from nouns in terms of case inflection is that the former inflect overtly for plural oblique, whereas the latter do not: cf. *altor prieteni* 'of/to other friends', not *altor* ***prietenor*. We must therefore ensure that *-or* as the exponent of the plural oblique case cannot be attached to word-forms, the lexical entries and CLRs of which do not include MP, but only include N.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

As we already know, plural and oblique are the non-default values of number and case respectively in Romanian. Case in the present framework manifests the insertion into the CLR of a feature CASE, possibly in combination with the other features ordered after the root. Just like NUM only appears in the CLR when its value is non-default, i.e. plural, CASE is inserted only if the word-form realizing the CLR is in the oblique case. The *-or* ending thus constitutes the exponent of the feature set {MP NUM CASE}, hence the CLR for *altor* 'of/to other (NPs)':

(49) {$_W$ ⟨$_\Sigma$ ⟨$_\Re$ alt⟩{MP NUM CASE}⟩ {N}}

{N} is unordered as it should be since *altor* is common to F and non-F (see [44]), so we may assume N bears no value in this form. (I keep MP in the RES rather than dumping it with N for reasons that will become clear later on).

Neither CASE nor NUM are proper to *altor*, but they come to it through agreement: cf. *casa altor prieteni* 'the house of other friends'. Given this, I assume the head noun in *unor prieteni* contains CASE, as it must in order to trigger agreement, hence CLRs (50)-(51) for *prieteni* and plural *fete* in oblique contexts :

(50) {$_W$ ⟨$_\Sigma$ ⟨$_\Re$ p r i e t e n⟩{NUM}⟩{N CASE}}
(51) {$_W$ ⟨$_\Sigma$ ⟨$_\Re$ fat⟩{ N:Fa NUM}⟩{CASE}}

Nouns thus differ from nominals in that CASE combines with NUM in the latter (see 49), whereas it never does in the former. In (50) CASE is unordered along with {N}; it is unordered by itself in (51) where {N:Fa} combines with NUM.

The point seems to be that not all feature collections are realizable. By collection, I mean a grouping of features in a given CLR that need not pertain to the same set. In particular, the feature collection N(:F) NUM CASE cannot be realized, irrespective of whether N is valued or not, therefore ordered or not (compare [50] with [51]). There is simply no exponent, wheras there is one, *-or*, for the collection (N MP NUM CASE) (see [49]).

The same account applies to the existence of non-F singular oblique *altui* where *-ui* realizes the collection (N MP CASE), {N} unordered, as opposed to the impossibility of ***prietenui*, where *-ui* would have to realize (N CASE), {N} unordered.[49]

Of course, separate endings for nouns and for nominals are banal (cf. Latin *ill**ius** amic**i*** 'of that friend'), and I am well aware that, to a large extent, the present account boils down to the statement that Romanian nominals do not inflect like nouns do. The down-boiling is not complete, however, for two reasons. First,

[49] The notion of collection does not run counter to the assumption that only ordered features are realized. Recall that unordered features in CLRs, although 'silent', are morphologically active. What we have to say, therefore, is that *-ui* in *altui* is the exponent of CASE and MP, in the context of bare N. In fact, collections heighten the realizational, non-incremental, character of the framework.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

Romanian is still peculiar insofar as, instead of having two distinct endings like Latin – or one ending fitting both classes as in *ill**orum** amic**orum*** 'of those friends' –, it has one ending to express plural oblique that goes with one class and excludes the other. Secondly, all I ever attempted was to formalize this state of affairs, not to 'explain' it, if there is such a thing. Yet, the formalization is not vacuous, because it will allow us to give meaningful answers to questions (c) and (d): Why do only F nouns overtly inflect for case? And why do they inflect as they do? Meanwhile, let us turn to question (b).

8.2. Do non-F nouns include CASE?

It seems we already answered this question in the foregoing subsection, as agreement facts led us to assume that all nouns include CASE even when it does not show up. What I wish to point out in the present subsection is that there may be more evidence bearing on this issue.

It is a peculiar feature of Romanian that unmodified definite nouns are **not** articulated when governed by prepositions that assign direct case (or perhaps no case at all), as shown in (52):

(52) pe scaun(*-ul) (*pe scaun*)
on chair(*-the$_{SG.NON\text{-}F.DIR}$)
on the chair

In the same context, nouns modified by attributive adjectives or any kind of complement are always articulated : cf. *pe scaunul cel mare* 'on the big chair', not **pe scaun cel mare*.[50] And note that 'on a (big) chair' would be *pe un scaun (mare)*, so there is no ambiguity as regards definiteness in (52).

In contrast, definite nouns governed by one of the few prepositions that assign oblique case **must** be articulated:

(53) deasupra scaunu*(-lui) (*deasupra scaunului*)
over chair-*(-the$_{SG.NON\text{-}F.OBL}$)
over the chair

Since ambiguity avoidance cannot be invoked – (52) not being ambiguous as we just saw – one way to make sense of these data is to assume that articulation is

[50] *Cel* is the so-called 'adjectival article' (for which see Pană Dindelegan 2003, Chapter 2). In traditional grammar, prepositions like *pe* 'on' – the overwhelming majority – are said to govern the direct case. In the present framework, they rather govern no case, i.e. they do not require CASE to be inserted in the CLR of the complement noun. There are two unexplained exceptions to the 'rule' exemplified in (52), namely *cu* 'with' (cf. *cu scaunul* 'with the chair') and the complex preposition *de-a* as in *a juca de-a baba-oarba* lit. 'to play the blind old woman', i.e. 'to play blind man's buff'.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

required in (53) in order to display the non-default case the non-F noun itself cannot overtly express. The implication, then, is that *scaun* does include CASE in (53), and that it shows through article concord.[51]

8.3. Why do only F nouns overtly inflect for case?

The assumption I wish to uphold is that CASE, when present, i.e. non-default, does not combine directly with the root, but with the N feature. It follows that CASE, albeit present in the CLR, is only expressed insofar as N also is, that is insofar as it too has a non-default value, that is F.

In non-F nouns, therefore, CASE is unordered along with bare {N}, showing only through concord or agreement:

(54) ⟨$_{W'}${$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\text{Я}}$ prieten⟩{ }⟩{N CASE}}{D CASE}⟩ *prietenului* 'of/to the friend'

In the set {D CASE}, realized as *-lui*, CASE is inserted as a consequence of concord with unrealizable {N CASE}.

In F singular nouns, in contrast, {N:F} is ordered in the stem list, so CASE can be expressed next to the root, and D in the extended word list, concords with it for g-gender and case (number being still default):

(55) ⟨$_{W'}$⟨$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\text{Я}}$ fat⟩{N:Fa CASE}⟩⟩{D:F CASE}⟩ *fet-e-i* 'of/to the girl'

{N:Fa CASE} is realized as *-e*, {D:F CASE} as *-i*. I give below the paradigm of the enclitic definite article combined with N, NUM, and CASE:

(56)

	N		N:F	
		NUM		NUM
	-(u)l	-i	-a	-le
CASE	-lui	-lor	-i	-lor

In the plural, as we know, CASE cannot enter the set {N(:Fa/e) NUM}, so again it only shows through concord and/or agreement: cf. *fet-e-lor* 'of/to the girls' where *-e* realizes {N:Fa NUM}, and *-lor* is the exponent of {D CASE}.

As mentioned in section 2, *-e* is not the only exponent of {N:Fa CASE}, but it is realized as *-i* in nouns like *pisică* 'cat', hence *unei pisici* 'of/to a cat', *pisicii* 'of/to the cat'. The allotment of Fa nouns to one or the other subclass is unpredictable. Yet, a closer look at the *-ă*/*-ĭ* subclass may prove fruitful.

[51] Of course, we still have no account of (52) itself, probably a separate phenomenon as suggested by the fact that F nouns, which can show CASE independently of articulation, behave alike in the same context: cf. *pe masă* 'on the table' (not **pe masa*) vs. *deasupra mesei* 'over the table' (not **deasupra mese*).

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

We already met with it in section 7, where we reached the conclusion that its distinctive property is being provided with different values of N in the singular and the plural, in accordance with generalization (32) : instead of N:Fa, the plural shows underspecified N:F, for which there is no exponent, hence the bare plural exponent *-ĭ* (see [39]). This is an account we can keep for CASE expression, assigning the following CLR to singular oblique *pisici* :

(57) {$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\Re}$ pisik⟩{CASE}⟩{N:F}}

Lone CASE would then receive the same exponent as lone NUM. Does (57) falsify the assumption that CASE steps in combined with N and is realized only if the latter is? I don't think it does, given the notion of collection introduced above. For features to belong to a collection rather than just sitting next to each other actually entails the same consequences than sharing the same set, insofar as a collection constitutes a discontinuous set in the sense that its elements occur in different subparts of the W set or list. We can thus maintain with (57) that CASE is in combination with N. But, one shall object, isn't it the case that N:F has no exponent, so CASE shouldn't be realized either ?

This is where we must qualify our account in an important way. We keep the notion that CASE in Romanian must share a feature collection or set with N. We also maintain that it is realized only if N is. Here is the qualification: if N is realized **in the paradigm the CLR is part of**, not necessarily in the same CLR. In non-F nouns, bare N is realized nowhere in the paradigm, so CASE has no realization whatsoever. In F nouns, in contrast, whatever subclass they belong to, N is always visible in at least one cell of the paradigm.[52]

Fe nouns forming their plural in *-ĭ* such as *carte* / *cărți* 'book(s)' (the majority) fall under the same account. That is to say, whereas direct singular *carte* has the CLR in (58), the oblique *cărți* is as in (59) :

(57) ⟨$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\Re}$ kart⟩{N:Fe}⟩⟩
(58) {$_{W}$ ⟨$_{\Sigma}$ ⟨$_{\Re}$ kart⟩{CASE}⟩{N:F}}

Case being thus expressed and interpretable in the 'minimal' (unarticulated) word-form of F nouns, independent syntactic factors are responsible for the ill-formedness of, e.g., **rochia fete* or **coada pisici* compared with *rochia unei fete* 'the dress of a girl', *coada unei pisici* 'the tail of a cat', *rochia fetei* 'the girl's dress', or *coada pisicii* 'the cat's tail'.

We have a problem, however, namely non-F nouns ending in *-e* such as *frate* / *frați* 'brother(s)'. Why don't they inflect for case, like Fa *pisică* or Fe *carte* 'book' (for which see below), since N is visible in them – cf. *iubirea unui frate*

[52] This account is probably equivalent to a rule of referral such as (12). Whether the relation is one of the notational variants is stuff for later research.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

(**fraţi*) ‘a brother’s love’?[53] I have to admit to failure in finding a fully satisfying solution. The best I can come up with in keeping with the previous assumptions is this: apparently, a **negative** value such as F–, the value of N we assigned to such nouns, although it is expressed, does not count as a trigger for CASE expression, contrary to positive values such as Fa or Fe. Recall that F– was assumed to represent a relative non-default value within the globally default value of N. Perhaps this is the crucial point: in order to allow for CASE realization, N must not only be realized too, its value must be absolutely non-default.

Another problem is raised by ambigeneric nouns: why don’t they inflect for case in the singular since N:F seems to be present in their paradigm, namely in the plural? The answer is that ambigeneric nouns actually function with **two** paradigms, one for the singular, one for the plural. Recall that a conjunct of two nouns with opposite genders is equivalent to one ambigeneric (see [43]). Of course, the two paradigms are connected inasmuch as they pertain to the same lexeme. This is exactly what I defended in section 7.

Bringing ambigeneric nouns into the picture shows that nouns have to be **entirely F** in order to inflect for case in the singular. This I call the **all-F condition**. Notice it is distinct from the paradigmatic condition that says that an F noun inflects even if N:F is not realized in all cells of the paradigm.

Although necessary for case inflection, F’s presence may not always be sufficient, however. I am referring to the repeatedly mentioned nonmatchers such as *aghiuţă* ‘little devil’, *paşă* ‘pasha’, etc. Their behaviour as far as case marking is concerned is complex (cf. Lombard, Gâdei 1981, II: 11-12). A first subcase is *tată* ‘father’, Fa in the singular (*tata* ‘the father’),[54] non-F in the plural (*taţii* ‘the fathers’), and which does not inflect in accordance with the all-F condition – cf. *iubirea unui tată* (**taţi*) ‘a father’s love’, *iubirea tatălui* ‘the father’s love’.[55] *Paşă* ‘pasha’ is like *tată*, except that the singular articulated form does inflect for case: cf. *smintirea paşei ~ paşii* / **paşalui* ‘the pasha’s insanity’ (*paşii* ‘the pashas’), whereas the unarticulated form does not: cf. *smintirea unui paşă* (**paşi*) ‘a pasha’s insanity’. This shows the paradigmatic and the all-F conditions to be indeed independent, in such a way that unarticulated *paşă* obeys the all-F condition as does *tată*, whereas articulated *paşa* behaves like ordinary F nouns. Why this should be so, however, is probably one of those historical accidents about which the present study has nothing to say. Nor can we hope to explain why *beşleagă* ‘old fogey’ presents us with two forms in the articulated singular oblique: *beşlegii* ‘of/to the old fogey’ similar to *paşei ~ paşii* (cf. *beşlegii* ‘the old fogeys’), and the curious mix *beşleagăi* combining the non-inflected form of not-all-F *tată* with the oblique form of the F article, *-i*.

[53] Compare *scrierea unei cărţi* (**carte*) ‘the writing of a book’.

[54] One also finds *tatăl* ‘the father’, which shows assimilation to non-F nouns despite the *-ă* ending.

[55] Recall we analysed such nouns as ‘reverse’ ambigenerics, showing F in the singular and non-F in the plural.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

8.4. Why do F nouns inflect as they do ?

This is question (d), and it divides up into two: (d') Why do we find the same exponent for the feature sets {N:F CASE} (singular oblique) and {N:F NUM}(plural)? (d'') Why is there no overt case distinction in the plural of nouns?

We already have an answer to (d''), namely that the only exponent available for the feature collection (N NUM CASE) requires that it should also include the feature MP, so *altor* whose CLR does include it, is a legitimate word-form, whereas **prietenor* or **fetor* are not. One might wish for a not so openly self-repeating account, however. A possible one would run as follows.

Let us keep the conclusion of the foregoing paragraph, interpreting it as meaning that CASE needs a pronominal feature to combine with in order to be expressed, but let us assume in addition that the RES of nouns is subject to a constraint that forbids it to contain more than two features. In other words, no cumulative exponent may lump together more than two features. The Romance languages at least seem to provide good empirical support for this hypothesis (see Kihm 2006 about Old French).

The constraint rules out **fetor* right away: in *fete* 'girls', the RES contains N:Fa and NUM, so there is no room left for the additional feature CASE. This feature may combine with D, however, because D is not in the RES and it does include a pronominal feature, hence *fetelor* 'of/to the girls', where *-e* means only 'plural' (see below). Singular oblique *fete* 'of/to (a) girl', in contrast, is provided with a RES that contains no more than N:F and CASE (see below), and N conceivably is pronominal like D.

In non-F plural *prieteni* 'friends', on the other hand, the RES only contains NUM, so the constraint would not oppose **prietenor*. What precludes it is the fact that NUM is not a pronominal feature, meaning that CASE cannot combine **only** with it. Neither can it combine with unordered bare N. The only possible source for the required pronominal feature is D, external to the 'minimal' word. In this way unarticulated substantives turn out to be uninflectable for case, unless they are F and singular.

Such a situation never occurs with nominals: in *altor*, for instance, the pronominal feature MP is present in the collection along with NUM and CASE (see [49]), thus allowing for the realization of *-or* to the immediate right of the root.

The answer to (d') involves the notion of toggle exponent proposed in the introduction, and the argument is the following. In languages like Romanian (at least) exponents express non-default feature values. Ideally, therefore, there should be as many exponents as there are non-default feature values. It is seldom so, however, and it is certainly not the case in Romanian where nouns have two non-default values to express for the features NUM and CASE, but only one exponent is available, namely *-e*/*i*. This lone exponent is then recruited to express either one of the two non-default values, that is CASE:oblique when NUM's value is default

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

and not in need of being expressed, or NUM:plural when CASE's value is default. For independent reasons (see preceding paragraph) CASE cannot be expressed at all in plural nouns. This means that CASE, present in plural CLRs to account for concord and agreement facts (cf. *prietenilor* 'of/to the friends, *unor fete* 'of/to some girls'), is unordered, so only NUM (in *prieteni*, *pisici*, or *cărtsi*) or the set {N:F NUM}(in *fete*) is associated with an exponent, perforce the same as that of CASE when it is apt to be realized, that is when NUM's value is default and N is absolute non-default (i.e., N:Fa or N:Fe).

Romanian *e/i* is thus similar to Old French *-s*, which expresses non-default subject case when number is default: cf. *li murs* 'the wall' (subject) vs. *le mur* 'the wall' (object); and non-default number when case is default: cf. *les murs* 'the walls' (object) vs. *li mur* 'the walls' (subject) (see Kihm 2006). It is also similar to Classical Arabic *ta-* in the imperfective paradigm, which expresses non-default person (2) when gender and number are default, i.e. masculine singular: cf. *taktubu* 'you (will) write' (said to a male); and non-default gender (feminine) when person and number are default (3.SG): cf. *taktubu* 'she (will) write(s)'. Strange 'homonymies' are thus accounted for. It would be an interesting research programme to examine whether more such cases can be unearthed, as they certainly tell us something important about the economy of the language faculty, namely its capacity to squeeze maximal yield out of privative contrasts.

I wish to conclude this section with an apparently minor observation, yet one which brings further empirical support to the present analysis. I am referring to the fact that a number of F nouns which have no plural for semantic reasons – e.g., *foame* 'hunger', *sete* 'thirst', *linte* 'lentils', *lene* 'sloth' – also do not show a special form for the singular oblique: cf. *durerile foamei* 'the pains of (the) hunger', not ***fomii* as the form would look if it existed. That it is not simply due to the absence of an actual plural is shown by those other F nouns that do not normally pluralize, but do have a special singular oblique form because a plural form is at least virtually available. Such is the case with *pace* 'peace', for which *păci* is not used to mean 'peaces' but occurs in *unei păci* 'of/to a peace' (see Lombard 1974: 49-50).

A rule of referral such a (12) accounts nicely for this evidence: if the form of the singular oblique is set in reference to that of the plural, it is only to be expected that it cannot be found whenever the plural does not exist at all. Our toggle exponents – perhaps no more (and no less) than the embodiment of rules of referral – allow for an equally elegant account, however. For *-e/i* to be a toggle exponent indeed implies that it means non-default CASE **and** NUM in every paradigm that is such that both values must be expressed. The toggle exponent is therefore defused the moment either one of the two values is erased. This is the case with *foame*: being an Fe noun, it satisfies the conditions for double expression of CASE and NUM – except that NUM is forbidden to assume its non-default value. The toggle exponent for non-default CASE **and** NUM becomes unusable as a consequence. In nouns like *pace*, in contrast, non-default NUM is not in use, but native speakers know what the word-form including it would be like if they needed it.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

Nouns like *foame* are thus uninflectable. They should be distinguished from other seemingly uninflected nouns such as *învățătoare(a)* '(the) schoolmistress', i.e. F nouns ending in *-toare* and denoting persons (cf. *smintirea învățătoarei* 'the schoolmistress's insanity'). The problem with these nouns is not that they do not pluralize, but that they have the same form in the plural as in the singular, that is they belong to the minority group of Fe nouns that show the same plural ending *-e* as Fa nouns (cf. *învățătoarele* 'the schoolmistresses'). We may thus safely assume that, save for this phonological quirk, they behave in a fully normal way as far as case inflection is concerned.[56] Support for this conclusion comes from F nouns with the same ending but not referring to persons, e.g., *lipitoare* 'leech', which inflect unsurprisingly: *unei lipitori* 'of/to a leech', *lipitori* 'leeches'. Note that the systematic contrast between *învățătoare* and *lipitoare* (also *sărbătoare* 'festival') is a good argument – next to compelling syntactic ones – for considering that the encyclopaedic feature [female person], which triggers the interpretation of F g-gender as feminine w-gender, may be morphologically relevant, which in turn has consequences for how components are interfaced. I leave this for future research.

9. SUMMARY

We started with a putative rich system of three genders and four (or five) cases for Romanian substantives. At the close of our investigation, we end up with a very streamlined organization. Having defined the framework and the operative concepts, our first step was to draw a clear line between gender *qua* 'natural' classification (w-gender, where 'w' stands for 'world') and gender *qua* grammatical class (g-gender). This distinction allowed us to bring down the number of non-default values of the substantivizing feature N to one, namely F, with two subvalues associated with different exponents, Fa and Fe. Remark that contrary to F, which I take to be substantial, Fa and Fe are notational devices abbreviating the alternative realization rules (60) and (61):

(60) RR(a) $_{N:F}$ (<X, σ>) $=_{def}$ <X*ă*, σ>, X any root compatible with N:F
(61) RR(b) $_{N:F}$ (<X, σ>) $=_{def}$ <X*e*, σ>, X any root compatible with N:F

Likewise, F– should be seen as a diacritic used in order to mark off those cases where default N identifies a stem-internal site to the right of the root (the RES) instead of being unordered, being thus associated with an exponent *-e*. As we saw, N:F– does not contrast with bare N as far as case inflection is concerned. Two classes of nouns thus result: non-default F nouns and default non-F nouns. In the latter, N receives no value beyond substantiveness itself.[57]

[56] Note it is the suffix *-toare* that is in cause as shown by the normalcy of, e.g., *profesoară* / *profesoare* 'female professor(s)'.

[57] This is the only class in so-called 'genderless' languages such as Hungarian or Turkish.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

Non-F nouns denoting persons or personalized beings are interpreted as masculine in terms of w-gender; F nouns under the same conditions are feminine, except for a limited stock such as *paşă* 'pasha' or *popă* 'pope' which happen to refer to male characters and are thus treated as masculine. Nouns that do not denote persons or personalized beings are not assigned w-gender. Many of them, and a sizeable number of person-denoting nouns as well, receive the conventional label 'neuter', but this is not a self-standing category in any way, similar to the Latin or Slavic neuter. Such nouns are actually ambigeneric in the sense that they are non-F in the singular, but F in the plural. Ambigenericity is demonstrated by agreement patterns in distributive constructions and by the fact that predicate adjectives qualifying mixed-gender conjuncts appear in the plural F form (see [42] and [43]).

The g- vs. w-gender distinction has a syntactic effect, as it leads us to assume two different feature-matching processes, agreement and concord. Concord concerns g-gender, and it is local, from the noun to the article. The articulated form of, e.g., *paşă* is thus *paşa* 'the pasha', showing the F article concording with an F noun which is otherwise interpreted as masculine. Agreement, on the other hand, is sensitive to w-gender and it targets all the other items that covary with the noun, i.e. demonstratives, quantifiers (including the indefinite determiner) and attribute or predicate adjectives, hence *Acest paşă tânăr e smintit* 'This young pasha is crazy', where *acest* 'this', *tânăr* 'young', and *smintit* 'crazy' all show their non-F form, here interpreted as masculine in agreement with the nonmatching masculine w-gender of the F head noun. As it turns out, the contrast only shows up with this closed class of nonmatchers, which are moreover to be considered reverse ambigenerics, insofar as they revert to non-F-hood in the plural (cf. *paşii* 'the pashas', and see above for some quirks). It cannot appear as matter of principle with inanimate-denoting nouns which ignore w-gender.[58]

Plurality represents the meaning of the feature Number (NUM) when it enters a noun's CLR. Absence of NUM means singular by default. If N is unordered (having then no value or the underspecified value N:F), NUM solely identifies the RES and it is associated with the 'plain' number exponent *-ĭ*. Otherwise, NUM combines with N:Fa/e identifying the RES, and the cumulative exponents are *-e* (*case* 'houses') or *-urĭ* (*trenuri* 'trains').

Finally, case inflection (leaving the vocative aside) also boils down to the contrast of default direct case, actually equivalent to no case, vs. non-default oblique case, which it is tempting to call case *tout court*. Morphological case expresses the feature CASE. Only on the article and the determining and pronominal items I called nominals, containing the (avowedly *ad hoc*) feature MP, is this contrast always visible. On nouns, the evidence for case is the fact that F nouns (a) overtly inflect for case while non-F nouns don't; (b) show the same form in the singular oblique as they do in the plural, where case and no case are

[58] Event-referring *asta* 'this' as in (31) remains special, however.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

nondistinct. I accounted for (a) by assuming that, in order to be expressed, CASE must combine with expressed N, that is N with the value Fa/e. This effectively limit case exponence to F nouns; and it also limits it to the singular since in the plural either N:Fa/e combines with NUM, precluding all further combination because of the constraint on the RES's size, or NUM is the sole identifier of the RES and CASE cannot combine with it because it is not pronominal as N and D are. Moreover, it is not enough for a noun lexeme to be partially F, as ambigeneric nouns are, it must be 'all-F' to inflect for case. As for (b), it points to the conclusion that the overt exponent of case *-e* or *-ĭ* is provided with what I call a 'toggle' property, which makes it similar to Old French *-s*, that is an exponent of either one of two non-default values, but not of both at the same time.

In sum, Romanian noun morphology is seen to make maximal usage of privative contrasts in the sense of Trubetzkoy's *oppositions privatives*, which allows it to function with one non-default g-gender (F), one non-default number value (plural), one case (oblique), and one toggle exponent (with two forms) expressing either case or non-default number, never both, in all-F nouns only.

10. CONCLUSION: A FEW COMPARATIVE REMARKS

At the end of section 2 the question was raised: can a significant connection be found between the fact that only (all-)F nouns inflect for case and ambigenericity? Clearly, both crucially involve the feminine in traditional parlance, that is the non-default g-gender F. As we saw, ambigeneric nouns in their plural F form must 'know' they are non-F in the singular; and similarly their singular non-F form must 'know' the plural to be F. Another, especially clear example that proves it is the following, in which *două* 'two (eggs)' is F even though *un ou* 'an egg' is non-F because *ou* / *ouă* happens to be ambigeneric:

(62) Mi-a dat numai un ou, dar vream două.
to.me-has given only an $egg_{non\text{-}F}$ but I.wanted two_{F}
S/he gave me only an egg, but I wanted two.

In the same way, a singular F noun has to 'know' it is also F in the plural in order to inflect overtly for case – cf. *tată* 'father' above, which does not inflect because it is non-F in the plural. Put differently, ambigeneric nouns are not-all-F – singular non-F vs. plural F for the majority, the reverse for a closed class of nonmatchers in terms of the g-/w-gender correspondance – whereas inflectable nouns are all-F. In both cases it is thus a matter of having F in the paradigm, as both groups contrast with non-F nouns (traditional masculines) which do not overtly inflect for case and pluralize unremarkably.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

A question we should ask, therefore, is whether other languages that can be interestingly compared with Romanian also assign such a role to the feminine *qua* g-gender. A language that immediately comes to attention is Albanian, since it is related to Romanian by rather close areal bonds in addition to the more distant genetic kinship (cf. Sandfeld 1930), and its vocabulary also includes a large amount of ambigeneric nouns pairing singular non-F with plural F, as examplified in (62a) and (62b) (see Boissin 1975: 73):

(62a) mal -i (*mali*)
mountain-the$_{non\text{-}F}$
the mountain
(62b) mal -e -t (*malet*)
mountain-PL-the$_{F}$
the mountains

Yet, Albanian also possesses a small class of really neuter nouns, i.e. neuter in the sense of Latin or Slavic, and its noun inflection is more elaborate than that of Romanian. For this reason, a potentially more fruitful comparison, I think, is with Italian. In the following I will lean heavily upon Acquaviva's (2002) enlightening work in the matter.[59]

Italian is well-known for having a sizeable number of F plurals in *-a* paired with non-F singulars in *-o* such as *il braccio* / *le braccia* 'the arm(s)', *il cervello* / *le cervella* 'the brain(s)', *il muro* / *le mura* 'the wall(s)', *il uovo* / *le uova* 'the egg(s)', etc. Despite the etymology of the ending (Latin neuter plural *-a* as in *templum* / *templa* 'temple(s)'), such pairs cannot be directly compared with the Romanian ambigenerics, however. The difference is both semantic and morphosyntactic.

From the semantic viewpoint, Italian F plurals in *-a* are mostly collective plurals referring to wholes or masses, and as such they usually coexist with ordinary plurals in *-i* referring to pluralities of tokens. For instance, *le cervella* refers to the whole of someone's brainstuff (as in *farsi saltare le cervella* / **i cervelli* 'to blow one's brains out'), whereas *i cervelli* designates a plurality of separate brains, as in *la fuga dei cervelli* / **delle cervella* 'the brain drain'.[60] In some cases the 'collective' notion has to be qualified somewhat: for instance, *le braccia* refers to arms as body parts, whereas *i bracci* may denote arms of (old fashioned) record players, that is arms that do not necessarily go in pairs. In still

[59] I am much indebted to Paolo Acquaviva for sending me a partial draft of his forthcoming study of lexical pluralization.

[60] I am grateful to Lucia Tovena (p.c. 3/30/05) for this example. In Rumantsch this formation gave rise to an almost fully productive collective class whose members, contrary to (Modern) Italian, are actually singular: cf. *il crap* 'the stone', *ils craps* 'the stones', *la crappa* 'the (heap of) stones' (Liver 1982: 24). Note that Romanian has one and only one plural in *-ă*, viz. *ou(l)* / *ouă(le)* '(the) egg(s)', which happens to translate Italian *il uovo* / *le uova*.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

other cases there is no alternative 'regular' plural – e.g., *il miglio* / *le miglia* 'the mile(s)' (**i migli*), *il riso* / *le risa* 'the laughter(s) (**i risi*), *il uovo* / *le uova* 'the egg(s)' (**gli uovi*), etc. – so the issue of a collective meaning of the *-a* plural may seem moot for these items. Yet, it is clear that Italian *-a* plurals always mean plural 'and something', even though defining the 'something' may not be a matter of course in a few cases.

Italian *-a* plurals, on the other hand, are morphosyntactically F as evidenced by distributive constructions such as *Le uova costano sessanta centesimi l'una* 'Eggs cost 60 cents each' (see Acquaviva 2002), in which non-F *l'uno* would be ungrammatical although singular *uovo* is non-F. This contrasts sharply with equivalent Romanian constructions as in (42), (43), and (62).

Compared with Italian *-a* plurals, Romanian F plurals paired with non-F singulars therefore appear as mere plurals which can, but need not be interpreted as collectives or whatever it is that *-a* plurals have which ordinary plurals don't. One would then be tempted to discard any connection between the two phenomena. There may be one, however.

The feature set Italian *-a* expresses may be assumed to be {N:F GROUP}, taking GROUP as a rough label for a functional feature akin to NUM, but with partially different semantics (on these matters, see Link 1983; Ojeda 1992; Zabbal 2002). According to Acquaviva (2002), '*-a*-collectivization' is a derivational process, whereas ordinary pluralization is inflectional. This contrast of derivation vs. inflection is patently too crude, however. Pluralization pertains to inherent inflection, being thus indeed closer to derivation than to contextual inflection (e.g., case). Romanian ambigenerics make this especially obvious, the plurals of which constitute parallel developments from the same root with respect to the corresponding singulars. In a pair such as *fir* / *fire* 'thread(s)', for instance, the CLR of the singular is (64), while that of the plural is (65):

(64) $\{_{W}\langle_{\Sigma}\langle_{\Re}$ fir$\rangle\{\ \}\rangle\{N\}\}$
(65) $\langle_{W}\langle_{\Sigma}\langle_{\Re}$ fir$\rangle\{$N:F NUM$\}\rangle\rangle$

This is similar to the CLRs we would assign to *(il) muro* 'the wall' (cf. [66]) as opposed to *(le) mura* 'the walls' (cf. [67]):

(66) $\langle_{W}\langle_{\Sigma}\langle_{\Re}$ mur$\rangle\{$N:o$\}\rangle\rangle$ [61]
(67) $\langle_{W}\langle_{\Sigma}\langle_{\Re}$ mur$\rangle\{$N:F GROUP$\}\rangle\rangle$

[61] Take N:o to be a mere label for non-F nouns ending in *-o* (as distinct from non-F nouns in *-e* such as *il bicchiere* 'the glass').

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

In both cases, the F value of N comes up as an inherent property of the plural or 'collective', which cannot be predicted from the singular. In that sense, Italian *-a*-plurals and Romanian ambigeneric nouns are indeed analogous to each other and, for instance, to Arabic 'broken' plurals (see Kihm 2006).

There is therefore a breach of lexical integrity, meaning the notion that in a paradigm A_i, A_j, A_k…, the base A should be invariant in terms of feature composition, in particular it should belong to the same gender or word class. On the other hand, the element of plurality in Italian *-a*-plurals is GROUP, which is more specific than NUM, for it implies it, but not the reverse. This is why *-a*-plurals normally cannot denote individual plurality, hence the parallel existence of 'ordinary' plurals, with a few exceptions. In Romanian, in contrast, the plurality feature is simply NUM – as it is in Arabic broken plural, despite the speculative origin of the latter as collectives, as opposed to the suffixal 'sound' and 'real' plurals. In fact, it can be shown that the difference between broken and sound plurals is purely morphological, synchronically at least (see Kihm 2003, 2006).

Such is the case in Romanian as well, where the counterparts of Italian ordinary plurals and of Arabic sound plurals are the plurals of non-F nouns (*prieteni* 'friends') and of all-F nouns (*fete* 'girls', *pisici* 'cats'). The Romanian system thus consists in two orthogonal contrasts: as far as inflection is concerned, all-F nouns, which inflect for case-or-number, contrast with non-F and ambigeneric nouns, which do not; in terms of lexical integrity, on the other hand, ambigeneric nouns, which add F in the plural, contrast with non-F and all-F nouns, which add nothing but plurality. There is therefore a partial correlation between non-inflection and lack of lexical integrity, since ambigeneric nouns, despite taking F, or rather because they take it only in the plural, do not inflect. Why should that be so? And why is F the intrusive value?

Let us once again assume pluralization to be a near-derivational process. Derivation is not concerned with lexical integrity. Actually, its common effect is to change the category of the base it applies to. We therefore expect pluralization to have that capacity as well, albeit in a lesser measure, since inherent inflection stands so to speak halfway between real derivation and real inflection. Romanian ambigenerics and Arabic broken plurals show this quite plainly. Yet, plural formation is a function from nouns to nouns. Now, two features only, it seems, can be changed in a noun without erasing its identity as a noun: its class or gender, i.e. the value of N combined with the root, and its type as a count or a mass noun. Italian *-a*-collectivization effectuates both changes: *cervella* is mass and F, whereas *cervelli* is count and non-F, both being plural, group or individual. In Romanian ambigenerics, only the first change takes place.[62] What matters, however, is that

[62] It is a matter for discussion whether only the first or both changes occur with Arabic broken plurals. I leave this issue aside.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

such changes are entirely in the order of things given the near-derivational nature of plural formation. Why they are not more frequent cross-linguistically is an interesting but, I think, subordinate question.

Finally, all evidence points to the fact that the alternations are oriented: they go from the default to the non-default. Mass or group plural is likely to be relatively non-default with respect to count or individual plural, themselves absolutely non-default with respect to singular. On the other hand, in a class system like that of Romanian based on the privative contrast of something vs. nothing, the morphologically marked g-gender is *ipso facto* non-default. We therefore predict there cannot exist a language like Romanian except that it would systematically pair F singular with non-F plural. Whether this prediction is borne out is something that must be left for future research. Let me just underline that the adverb 'systematically' of the preceding sentence is crucial. Indeed, nouns like *paşă* / *paşi* 'pasha(s)', which seem to belie the prediction, are exceptions as we saw. Moreover, their singular F-hood is partial since it only triggers local agreement, whereas concord is non-F. 'Gender polarity' languages, on the other hand, that is languages that always show opposite genders in the singular and the plural, the stock example of which is Somali, are **not** like Romanian. In Somali 'masculine' and 'feminine' seem to be equally non-default, which suggests that what Somali shows is a very impoverished noun class system rather than g-gender (see Kihm 2005).

REFERENCES

Gramatica limbii române (GA), 1963, vol. I, Bucureşti, Editura Academiei.
Gramatica limbii române (AR), 2005, vol. I, Bucureşti, Editura Academiei Române.
Ackema, P., A. Neeleman, 2004, *Beyond Morphology: Interface Conditions on Word Formation*. Oxford, Oxford University Press.
Acquaviva, P., 2002, "Il plurale italiano in *-a* come derivazione lessicale", *Lingue e linguaggio,* 2, 295–326.
Aikhenvald, A. Y., 2000, *Classifiers : A Typology of Noun Categorization Devices*, Oxford, Oxford University Press.
Anderson, S. R., 1992, *A-morphous Morphology*. Cambridge, Cambridge University Press.
Aronoff, M., 1994, *Morphology by Itself: Stems and Inflectional Classes*, Cambridge, MA: MIT Press.
Beard, R., 1995, *Lexeme-Morpheme Base Morphology: A General Theory of Inflection and Word Formation*, Albany, State University of New York Press.
Blevins, J. P., 2005, "Remarks on gerunds", in: C. Orhan Orgun, P. Sells (eds), *Morphology and the Web of Grammar. Essays in Memory of Steven G. Lapointe*, Stanford, CSLI Publications, 25–47.
Boissin, H., 1975, *Grammaire de l'albanais moderne*, Paris, Chez l'auteur.
Booij, G., 1996, "Inherent vs. contextual inflection and the split morphology hypothesis", in: G. Booij, J. van Marle (eds), *Yearbook of Morphology 1995*, Dordrecht, Kluwer, 1–16.
Bourciez, E., 1910/1967, *Eléments de linguistique romane*, Paris, Klincksieck.
Brâncuş, G., 2004, *Introducere în istoria limbii române* [*Introduction to the History of the Romanian Language*], Bucureşti, Editura Fundației România de Mâine.
Corbett, G. G., 1991, *Gender*, Cambridge, Cambridge University Press.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

Corbett, G. G., 1998, "Morphology and agreement", in: A. Spencer, A. M. Zwicky (eds), *The Handbook of Morphology*, Oxford, Blackwell, 191–205.
Corbett, G. G., 2005, "The canonical approach in morphology", in: Z. Frajzynger, A. Hodges, D.S. Rood (eds), *Linguistic Diversity and Language Theories*, Amsterdam, John Benjamins, 25–49.
Cornilescu, A., 1992, "Remarks on the determiner system of Romanian: the demonstratives *al* and *cel*", *Probus,* 4, 189–260.
Croitor Balaciu, B., 2004, "Vocativul în româna actuală" [The vocative in contemporary Romanian], in: Gabriela Pană Dindelegan (ed.), *Tradiţie şi inovaţie în studiul limbii române* [*Traditions and Innovations in the Study of the Romanian Language*], Bucureşti, Editura Universităţii din Bucureşti, 73–78.
Daniliuc, L., R. Daniliuc, 2000, *Descriptive Romanian Grammar: An Outline*, München, Lincom Europa.
Dobrovie-Sorin, C., 2000, "(In)definiteness spread: from Romanian genitives to Hebrew construct states nominals", in: V. Motapanyane (ed.), *Comparative Studies in Romanian Syntax*, Amsterdam, John Benjamins.
Farkas, D., 1990, "Two cases of underspecification in morphology", *Linguistic Inquiry*, 21, 539–550.
Guerssel, M., J. Lowenstamm,1990, *The derivational morphology of the Classical Arabic verb,* Unpublished ms., Université du Québec à Montréal, Université Paris 7.
Halle, M., 1973, "Prolegomena to a theory of word-formation", *Linguistic Inquiry*, 4, 3–16.
Halle, M., 2000, "Distributed morphology : impoverishment and fission", in: J. Lecarme, J. Lowenstamm, U. Shlonsky (eds), *Research in Afroasiatic Grammar*, Amsterdam, John Benjamins, 125–149.
Halle, M., A. Marantz, 1993, "Distributed morphology and the pieces of inflection", in: K. Hale, S. J. Keyser (eds), *The View From Building, 20: Essays in Linguistics in Honor of Sylvain Bromberger*, Cambridge, MA: MIT Press, 111–176.
Harris, J., 1999, "Nasal depalatization *no*, morphological wellformedness *sí*. The Structure of Spanish Word Classes", in: Karlos Arregi *et al.* (eds), *Papers on Morphology and Syntax, Cycle One*. *MIT Working Papers in Linguistics*, 33, 47–82.
Jackendoff, R., 1997, *The Architecture of the Language Faculty*, Cambridge, MA: MIT Press.
Jackendoff, R., 2002, *Foundations of Language: Brain, Meaning, Grammar, Evolution*, Oxford, Oxford University Press.
Janda, R. D., 1983, "Morphemes aren't something that grows on trees : morphology as more the phonology than the syntax of words", in: J. F. Richardson, M. Marks, A. Chukerman (eds), *Papers from the Parasession on the Interplay of Phonology, Morphology, and Syntax*, Chicago Linguistic Society, 79–95.
Kihm, A., 2003, "Les pluriels internes de l'arabe: système et conséquences pour l'architecture de la grammaire", *Recherches linguistiques de Vincennes*, 32, 109–155.
Kihm, A., 2005, "Noun classes, gender, and the lexicon-syntax-morphology interfaces: a comparative study of Niger-Congo and Romance languages", in: G. Cinque, R. S. Kayne (eds), *The Oxford Handbook of Comparative Syntax,* Oxford, Oxford University Press, 459–512.
Kihm, A., 2006, "Nonsegmental concatenation: a study of Classical Arabic broken plurals and verbal nouns", *Morphology,* 16, 69–105.
Klausenburger, J., 2001, *Coursebook in Romance Linguistics*, München, Lincom Europa.
Lever, R., 1982, *Manuel pratique de romanche : sursilvain – vallader*, Chur, Verlag Lia Rumantscha / Ligia Romontscha.
Link, G., 1983, "The logical analysis of plural and mass terms : a lattice-theoretical approach", in: R. Bäuerle *et al.* (eds), *Meaning, Use, and Interpretation in Language*, Berlin, Walter de Gruyter, 302–323.
Lombard, A., 1974, *La langue roumaine: une présentation*, Paris, Klincksieck.
Lombard, A., C. Gâdei, 1981, *Dictionnaire morphologique de la langue roumaine*, Bucureşti, Editura Academiei.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)

Lowenstamm, J., 1996, "CV as the only syllable type", in: J. Durand, B. Laks (eds), *Current Trends in Phonology : Models and Methods*, Oxford, Oxford University Press, 419–441.
Lowenstamm, J. (to appear), "On n, √, and types of nouns", in: J. Hartmann, V. Hegedus, H. van Riemsdijk (eds), *Sounds of Silence*, Oxford, Elsevier.
Lumsden, J. S., 1992, "Underspecification in grammatical and natural gender", *Linguistic Inquiry*, 23, 469–486.
Mallinson, G., 1984, "Problems, pseudo-problems and hard evidence – another look at the Romanian neuter", *Folia Linguistica*, 18, 439–451.
Mallinson, G., 1988, "Romanian", in: M. Harris, N. Vincent (eds), *The Romance Languages,* London, Routledge, 391–419.
Marantz, A. P., 1997, "No escape from syntax : Don't try morphological analysis in the privacy of your own lexicon", in: A. Dimitriadis *et al.* (eds), *Proceedings of the 21st Penn Linguistics Colloquium*, Philadelphia, UPenn Working Papers in Linguistics, 201–225.
Matthews, P. H., 1972, *Inflectional Morphology: A Theoretical Study Based on Aspects of Latin Verb Conjugation*, Cambridge, Cambridge University Press.
Matthews, P. H., 1974, *Morphology*, Cambridge, Cambridge University Press.
Ojeda, A. E., 1992, "The semantics of number in Arabic", in: C. Barker, D. Dowty (eds), *SALT II. Proceedings of the Second Conference on Semantics and Linguistic Theory*, Columbus, Ohio State University, 303–325.
Pană Dindelegan, G., 2003, *Elemente de gramatică: dificultăți, controverse, noi interpretări* [*Elements of Grammar : Difficulties, Controversies, New Interpretations*], Bucureşti, Humanitas Educațional.
Pollard, C., I. A. Sag, 1994, *Head-Driven Phrase Structure Grammar*, Chicago, The University of Chicago Press.
Ratcliffe, R. R., 1998, *The 'Broken' Plural Problem in Arabic and Comparative Semitic : Allomorphy and Analogy in Non-Concatenative Morphology*, Amsterdam, John Benjamins.
Sadock, J. M., 1991, *Autolexical Syntax : A Theory of Parallel Grammatical Representations*, Chicago, The University of Chicago Press.
Sag Ivan A., T. Wasow, 1999, *Syntactic Theory: A Formal Introduction*, Stanford, CSLI Publications.
Sandfeld, K., 1930, *Linguistique balkanique: problèmes et résultats*, Paris, Klincksieck.
Stump, G. T., 1998, "Inflection", in: A. Spencer, A. M. Zwicky (eds), *The Handbook of Morphology*, Oxford, Blackwell, 13–43.
Stump, G. T., 2001, *Inflectional Morphology: A Theory of Paradigm Structure*, Cambridge, Cambridge University Press.
Trubetzkoy, N. S., 1939, "Grundzüge der Phonologie", *Travaux du Cercle linguistique de Prague*, 7. [1969, *Principles of Phonology*, transl. C. Baltaxe, Berkeley, University of California Press]
Wechsler, S., L. Zlatić, 2003, *The Many Faces of Agreement*, Stanford, CSLI Publications.
Williams, E., 1981, "On the notions 'lexically related' and 'head of a word'", *Linguistic Inquiry*, 12, 245–274.
Zabbal, Y., 2002, *The semantics of number in the Arabic noun phrase*, Unpublished Master Thesis, University of Calgary.

BDD-A278 © 2007 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.34 (2026-11-13 23:56:30 UTC)