ROMANIAN DEFINITE ARTICLE REVISITED

Introduction

I shall attempt to resume a long, almost endless discussion: the origin of the Romanian definite article. Any grammar of Romanian or any comparative grammar the Romance languages (e.g. Tagliavini 1977) always observes that Romanian, an isolated case in the Romance family, has an agglutinated definite article. The typology is not indeed rare: Bulgarian, Albanian, Armenian, Basque and Swedish witness the same mechanism. We cannot approach the topic by analysing all these languages, yet a comparative analysis would be finally useful. In our case, it is obvious that Romanian cannot be isolated from Albanian and Bulgarian. A potential solution must explain the situation in ALL these three “Balkanic” languages, even if Romanian is not Balkanic stricto sensu.¹

The paper shall focus on the deep roots of the Romanian and Albanian definite article, its typological relations with other linguistic areas, and shall attempt to explain this isolated situation in the field of Romance linguistics. For sure, the Romanian definite article mainly reflects the Latin heritage. Nevertheless, by saying only this, the tableau is not complete: some forms are not Latin but Pre-Latin, Thracian. This paper will try to substantiate this assertion.

The definite article of nouns and adjectives

Though the facts are well known, I shall resume the basic facts and point out less known details. At a first glance things are so simple: the definite article reflects the agglutinated Latin demonstrative. And yet...

¹ The term Balkanhalbinsel ‘Balkanic Peninsula’ was coined in the year 1808 by the Berlin geographer Johann August Zeune starting from the Turkish word balkan ‘mountainous rocky land’ and presumably reflects a calque after Bulg. Stara planina. The word Balkan(s) had a tremendous success, especially in its extended meaning (including its political connotations). The original meaning was purely geographic and referred to the modern states of Bulgaria, Macedonia, Albania and Greece.
Sing. | Pl.
--- | ---
N. A. | -u, -ul, -le* | -a | -i | -le
G. D. | -(e)i, -(i)i | -lor † | -lor
V. | -ul-e ‡ | -

* The form -u is always colloquial and is attested in the oldest Romanian texts, whereas -ul is bookish. Both are used with the former Latin nouns of second declension or assimilated with them (e. g. lupus > Rom. lup). The form -le is both colloquial and bookish and is used with former nouns of third declension (e. g. canis, Acc. canem > Rom. ciine).

Therefore the definite forms are, e. g., lup-lupul, but ciine-ciinele.

† Lu is always colloquial, lui (with i pronounced as semivowel y) belongs almost exclusively to the written language. Identical to the personal pronoun, genitive-dative.

‡ Identical to the personal pronoun plural, genitive-dative.

Common nouns like om - omule, fecior ‘son’ – feciorule, but copil ‘child’ – copile.
The pattern is therefore: sometimes noun + definite article -ul + ending e (< Latin vocative -e), sometimes the definite article is not required (no rule). See below the case of personal names.

Masc. sing. N. A. -u is as old as the form -ul, despite the largely spread hypothesis that the colloquial form -u would be simplified from -ul. Oldest Romanian texts witness -u rather than -ul. It is true that the form -ul is the only accepted in written texts, whereas -u belongs to the spoken language and is in fact the unique spoken form. Masc. sg. N. A. form -u has an identical parallel in Albanian, e. g. shok – shok-u ‘a colleague, comrade’, zog – zogu ‘a bird’, etc. Useless to say that the form -u cannot be explained from Latin like all the other forms in Albanian, where the paradigms are more complicated.

Let us compare the Albanian forms:

The Albanian Definite Article

Forms in the Nominative singular
Masc. | -i or -u
Fem. | -a
N. | -t, -të
Pl. | -t, -të

72
<table>
<thead>
<tr>
<th>Type I</th>
<th>Type II</th>
<th>Type III</th>
<th>Type IV</th>
<th>Type V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ind.</td>
<td>Def.</td>
<td>Ind.</td>
<td>Def.</td>
<td>Ind.</td>
</tr>
<tr>
<td>N</td>
<td>-i</td>
<td>-u</td>
<td>-a</td>
<td>-t, tê</td>
</tr>
<tr>
<td>G</td>
<td>i</td>
<td>it</td>
<td>u</td>
<td>it</td>
</tr>
<tr>
<td>D</td>
<td>i</td>
<td>it</td>
<td>u</td>
<td>it</td>
</tr>
<tr>
<td>Ac.</td>
<td>in, nê</td>
<td>un, nê</td>
<td>n, nê</td>
<td>t, tê</td>
</tr>
<tr>
<td>Abl.</td>
<td>i</td>
<td>it</td>
<td>u</td>
<td>it</td>
</tr>
</tbody>
</table>

| Includes most masculine names. |
| Includes some masculine forms which end in -g, -k, -h or in a stressed vowel. |
| Feminine; includes also a few personal masculine names which end in -ê or -o, like tatê, tata. |
| Includes all neuter forms. |
| Plural forms. |

First conclusions: (1) form -u is as old as the form -ul, perhaps even older (see below); it is identical to Albanian paradigm in -u; (2) form -ul reflects indeed Latin ille with the link vowel -u-.

- Masc. sing. N. A. form -le is used in case of words ending in -e e. g. ciine < Lat. cane(m), i. e. former Latin names of third declension or assimilated to it. It obviously reflects Lat. ille.
- Fem. sing. N. A. -a reflects Latin illa. Things are more complicated with the G. D. form -ei and sometimes -ii. It is well known that Fem. sing. G. D. forms are identical to Fem. pl. N. A. forms (non-articled), e. g. fatâ ‘girl’ – fete which is both G. D. sing. and N. A. pl. The definite article is called to clarify the case: fetei ‘to the girl’ – fetelor ‘to the girls’. So we may question whether the article of the fem. sing. G. D. is -i, identical to Albanian -i, or -ei as hypothesised by many linguists, it is true not taking into account the Albanian forms. Difficult to decide: if the old paradigm in Romanian for fem. sing. is (indefinite v. definite respectively) -ǎ v. -e (e. g. fatâ – fete), then the definite article for fem. gen. is INDEED -i (pronounced as a semivowel y, not -ei, pronounced ey). It is true that the corresponding personal pronoun G. D. is ei (v. masc. lui), but for both genders the short (unstressed) form is i. I would not dare to derive it from Latin as many other linguists do.
- Fem. pl. N. A. -le reflects indeed the Latin demonstrative. The same is valid for the forms -lui and -lor identical to the oblique cases of the personal pronoun el (masc. sing.) – G. D. lui and ei (masc. pl.) – lor.
- Romanian neuter forms follow the general rule: masculine forms are used for the singular and feminine forms for the plural. Romanian neuter is therefore strictly different from Slavic or German neuter.

A brief survey with the corresponding example is perhaps useful:
Masculine (indefinite / definite)

Sing.

N. A.  om / om-u, om-ul  căine / căine-le  
G. D.  om / om-u-lui  căine / căine-lui

Pl.

N. A.  oameni / oameni-i  căini / căini-i  
G. D.  oameni / oameni-lor  căini / căini-lor

Note: The graphic sequence ii includes (1) the mark for plural -i + (2) the definite article -i; it is pronounced as vowel i, against the indefinite plural form in -i which is pronounced as a very short i, in fact a palatalisation of the previous consonant. Therefore, the pl. indefinite form lupi is pronounced /lupi/, whereas the pl. definite form lupii is pronounced /lupil proper. The various pronunciations of graphic i represent a hard try for the foreigners who study Romanian.

Feminine (definite / indefinite)

Sing.

N. A.  fat-ii / fat-a  femei-e / femei-a  
G. D.  fete-i / fete-i  femei / femei-i (i-i is pronounced /i/)  

Pl.

N. A.  fete / fete-le  femei / femei-le  
G. D.  fete / fete-lor  femei / femei-lor

Note: In feminine singular, the opposition definite – indefinite of the first class (ending in -ă) is the opposition ă/a. ă is the neuter vowel ă, usually the quality of a in unstressed position; a similar sound is Alb. ă and Bulg. ă. Again, the spelling ii reflects a normal /i/, against the spelling i which in Romanian usually reflects the semivowel /y/.

Romanian neuter has no special forms. It uses the masculine forms in the singular, and the feminine forms in the plural, with (sometimes) the mark of plural -uri which is only for neuter plural. Otherwise the paradigm follows the same rule: the neuter is masculine in the singular and feminine in the plural.

The Latin ille, illa, illud has been considered and accepted as the origin of the Romanian definite article. It is indeed so, yet NOT ALL THE FORMS reflect this origin. And, if we refer to Albanian, with which Romanian has indeed much in common, we can realise that the situation is not so simple. It is not simple indeed even if we ignore the Albanian forms, as usual with most linguists who have analysed the topic.

The popular Latin forms which explain the definite article not only in Romanian, but also in Western Romance languages, are:
This reconstruction (Ivănescu 1980: 133-134), one of the best I have knowledge of, does not explain all the Romanian forms, among these the feminine singular form -a and accusative sing. of the feminine pronoun -o. Ivănescu justly notes, on the next page, that the agglutinated position of the Romanian definite article should be explained as a Thracian influence. He refers, of course, to the known fact that both the definite article and the adjective follow the rule “first the noun, then the article and/or adjective”. This is the usual form in Romanian, though dialectally some forms of the definite article are placed before the nouns and, from various stylistical reasons, the adjective may be placed before the noun. But the standard, also archaic, order is noun + definite article or noun + adjective. He does not dare mention that those forms difficult to explain via Latin may very well be inherited from the Thracian substratum, like – on the other hand – the colloquial and dialectal forms of the verb a fi ‘to be’: îs ‘I am, they are’, îi ‘he, she is’, which can NEVER be explained from Latin, but rather reflect a Thracian heritage.

The limited purpose of this paper does not allow me to extend the discussion. It is high time to pass to the next step:

The definite article of demonstratives and adverbs!

Romanian is different from any other Romance or Germanic language by using a specific form of the definite article. The situation has not been properly observed by most linguists and grammarians, so I shall concentrate on it.

Romanian uses an invariable definite article -a for both demonstrativa and adverbia. The invariable form -a should NOT be confused with the feminine definite article of singular forms. Many linguists, by not confusing the situation, simply claim that this specific -a is a particle. It is not a particle, it is a genuine and very clear definite article. Some examples may prove relevant.
Demonstrativa

The invariable definite article for demonstratives is used for masc. and fem. sing and masc. and feminine plural forms. It is always -a. Examples:

- *acest* 'this' (masc. sing.) – *acesta* 'this (plus the definite article)'; no equivalent translation. approx. 'this [person] I am talking about'.

  *Acest om este sărac* ‘this man is poor’
  *Acesta este săracă* ‘this (definite: man I am talking about) is poor’

- same construction is applied to similar forms like *acel* 'that' – *acela*

- *această* 'this (fem. sing.)' – *aceasta* 'this (plus the definite article)'; no equivalent translation. approx. 'this [woman] I am talking about'.

  *Această femeie este săracă* ‘this woman is poor’
  *Aceasta este săracă* ‘this (definite: woman I am talking about) is poor’

- *acești* 'these' (masc. pl.) – *aceștia* (same meaning plus definite article); no equivalent translation, approx. 'these [persons] I am talking about'.

- *acei* 'those' (masc. pl.) – *aceia*; identical construction as above.

- *aceste* (fem. pl.) – *acestea* 'these [women]' without and respectively with the definite article.

- *acestor* (G. D. plural forms for both masc. and fem.) 'to these [men or women] – *acestora*.

- *acelor* (G. D. plural forms for both masc. and fem.) 'to those [men or women] – *aceslora*.

Adverbia

Few adverbs very frequently used witness the same definite article -a. Examples:

- *ades* (from adj. des < Lat. densus) 'frequently' – *adesea* (with link-vowel -e which, given its position, is pronounced like a semivowel: e-o-a).

- *pururi* 'for ever, eternally' – *pururea* (with the same link-vowel -e); also in the construction de-a pururi – de-a pururea (same meaning, same parallel without and with definite article respectively)².

Two exceptional forms: tată/tata ‘father’ and popă/popa ‘a priest’

Tată 'father' is articulated tata (identical to Alb. tatē, tata), and popă ‘a priest’ is articulated popa. It is outstanding that these two exceptional forms have never been properly analysed, according to my available information. The origin of tată is, of course,

---

² Pururi was initially a noun, of neuter gender, *pur, pl. pur-uri, presumably of Thracian origin and having the meaning 'fire' i.e. 'eternal fire'. For the peculiar evolution of this meaning see Paliga 1992, reprinted in Paliga 1999.
Latin *tata, -ae* m., used in colloquial Latin (the modern English equivalent would be ‘dad, daddy’). The masculine gender of the Latin original is preserved in Romanian.

Things seem much more complicated with the form *popă* ‘a priest’. All the dictionaries and studies I have knowledge of (no exception) indicate that the origin is Slavic *popo*, not Latin *popa, -ae* (also a colloquial word) ‘a priest in charge with sacrifices’. Rom. *popă* is also exclusively colloquial (against the formal, official term *preot < Lat. presbiterum*). Indeed the Slavic form *popo* cannot be avoided, nevertheless things are not so simple, because Slavic *popu* cannot result in Rom. *popă*. This origin is to be identified in NP Pop, against Popa. The only argument I have heard (never read) is that Lat. *popa* should have resulted in Rom. *poapă*. I doubt that such an evolution is possible, because (1) the diphtongation in the pre-final syllable (*o > oa, in literary Romanian, or *o > o, i. e. open short o, in regional Transylvanian Romanian*) is the EXCLUSIVE attribute of the femine gender, and (2) a Slavic masculine could NEVER result in a Romanian masculine noun with feminine aspect. In fact, beside *popă*4, there is only *tată*.5

It is impossible to accept the idea that Lat. *tata* and *popa*6, two colloquial Latin forms of masculine gender of the first declension, preserved in Romanian as *tată* and *popă* respectively use the definite feminine article. In these two forms, THE ONLY ACCEPTABLE HYPOTHESIS is that they preserve the archaic bi-gender (or bi-functional masculine-femine) article -a of Thracian origin. We cannot know the various paradigms of the Thracian noun, but it is safe and logical to assume that such an article did exist, as it has been preserved in some archaic Romanian forms belonging to the basic vocabulary.

A would-be form *poapă* is really impossible, as the diphtongation of the pre-final o in case of feminine words ending in -ă and (sometimes) -e is such a strong mark of the femine gender, that the rule is followed by the recent borrowings, e. g. director m. – *directoare*. *Popă* ‘priest’, with a deep mark of the masculine character, can never become *poapă* which sounds pejorative. The word is really sometimes heard with the meaning ‘a priest’s wife’.

---

3 Dr. Gheorghe Mihăilă, specialist in Old Church Slavonic and author of numerous books regarding the relations between the Romanians and the Slavs.

4 It is not the purpose of this paper to discuss the origin of Slavic *popo*, but I wonder whether the largely accepted theory which considers this word as reflecting Gr.-Lat. *papas* should not be rather replaced by a less comfortable theory, implied in this text, that it reflects Rom. *popă*. For further discussions regarding the oldest Romanian and Thracian borrowings in Slavic see Paliga 1996, passim.

5 Rom. *vodă, abridged from vojevoda*, also with feminine aspect, is an obsolete undeclinable form of Slavic origin. It is not used any more: the word disappeared from the common vocabulary when the historical and social context disappeared too. *Tată* and *popă* have remained words of the basic vocabulary.

6 Lat. *tata* belongs to the childish vocabulary, while the colloquial form *popa* is presumably of Etruscan origin (Ernout-Meillet 1959 s. v.).
The definite article of personal names

Personal names follow some other rules, i. e.:

- The G. D. form for masculine is placed before the noun, e. g. N. A. Petre - G. D. lui Petre. As always, the definite article is identical to the G. D. personal pronoun.

- The feminine personal nouns are ALWAYS articulated in the N. A. case: Ileana, Maria, also NL Sofia, Londra, unlike their masculine counterparts which are not. Masculine place-names follow the same rules as masculine common names (i. e. non-articled in N.A. basic forms).

- The feminine G. D. forms are identical to the common nouns. Nevertheless in contemporary Romanian the G. D. masculine form is used, though it sounds strangely: lu (instead of bookish lui) Ileana. The form is almost acceptable for foreign feminine personal nouns which cannot be included in a Romanian paradigm ending in -a or -e, e. g. N. A. Carmen – G. D. lui Carmen. The normal form would be *Carmenei, but it is merely theoretical: nobody uses it. Problems appear in written Romanian, not in colloquial Romanian.

But maybe the most interesting fact is represented by the personal family names ending in -a. They are represented by an important number of names of various roots: indigenous Thracian, Latin and Slavic. By tradition, a family name reflect the male ascendency, and some of them are indeed articulated with the masculine article -u (never ul, which is exclusively bookish), e. g. brad ‘a fir’ - NP Bradu, Brădeanu (indigenous Thracian root); lup (Latin lupus) – NP Lupu, Lupescu (Latin root), Mircea (from Sl. mir) etc.

Another category is represented by the forms ending in -a or -ea (a diphtong, with the specific semivowel e+, dialectally pronounced i+): Bradea, Lupea, Toma, Nicula, Mircea etc. As easily observable, a is the (masculine) definite article, not an -a ending as suggested by most scholars. If there are still doubts, I must add that in all such forms, the -a definite article may be replaced by the usual masculine definite article -u, without any change of sense, even though sometimes the -u-articled forms are not usual or never used as such.

Discussion

Romanian shares with Albanian and Bulgarian the specific agglutination of the definite article. Typologically this construction is also met in Swedish, Armenian, Basque and – according to recent theories – in Etruscan. By analysing all the available relevant data we can observe that:

- Romanian is closely related to Albanian in many aspects, specifically the masc. sing. form N.A. -u (colloquial in Romanian) and fem. sing. G. D. -i.
• Romanian is unique in preserving an invariable definite article -a for both adverbs and demonstratives. This form should not be confused with the fem. sing. form -a of presumably Latin origin. This article is also used in the case of two words belonging to the basic vocabulary: tată and popă, both of masculine gender.
• The other forms reflect Latin ille, illa and, for plural, the oblique cases of the personal pronouns are agglutinated and used as the definite article.
• The Bulgarian definite article is a calque after Thracian and/or Romanian. An accurate analysis could be made only after deciding whether Thracian was still spoken at the arrival of the first Slavic groups in the South Danubian region, which is very probable (a hypothesis well argumented by the Bulgarian School of Thracian Studies). It is feasible to admit that the agglutinated definite article in Bulgarian and Macedonian is a calque after the indigenous Thracian substratum, later consolidated under the (Proto-) Romanian influence.

References


Șăineanu, Lazăr 1929 (DU). Dicţionar universal al limbii române, ed. a 6-a, îngrijită de M. Stăureanu. Craiova.


Povzetek

NOV POGLED NA DOLOČNI ČLEN V ROMUNŠČINI

Določni člen v romunščini ni zgolj latinskega izvora, kot je mislila ogromna večina raziskovalcev, čeprav je večina oblik seveda latinskih. Za določni člen je namreč mogoče ugotoviti arhaične črte, nedvomno avtohtonega trako-dakijskega izvora. Zasledimo jih v obliki za moški spol ednine na -u, ki ima popoln ustreznik v albanščini, kakor tudi v določnem členu prislovov na -a, kar dela romunščino v romanski jezikovni družini za docela samosvojo. Doprinos avtohtonega trako-dakijskega jezikovnega fonda je treba tehtati v širšem kontekstu, v celovitosti vpliva substrata; ta je mnogo pomembnejši, kot je to pripravljena sprejeti večina raziskovalcev tega jezikovnega območja.