The Balkan Definite Article and Pseudo-Second Position
Author(s): Aaron Halpern

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via http://linguistics.berkeley.edu/bls/.

The Annual Proceedings of the Berkeley Linguistics Society is published online via eLanguage, the Linguistic Society of America's digital publishing platform.
The Balkan definite article and pseudo-second position
Aaron Halpern
Stanford University

1. Introduction

Lexical Phonology (Kiparsky 1982, Mohanan 1986) provides us with a theory of the output of the lexicon according to which all nonautomatic phonological and morphological processes are carried out in that component of the grammar. I will be assuming the essential correctness of this approach and indirectly addressing the issue of the relationship between the lexicon and the rest of the grammar. I want to emphasize that this assumption doesn’t a priori commit us to a Lexicalist view of syntax, though the proposal in this paper will be along those lines. Instead, this assumption, that the phonological aspects of word formation form a module, makes it easy to ask an important question:

If we define the output of the lexicon, in the Lexical Phonology sense, as the morphological word, what is the relationship between the morphological word and the syntax?

Current theories provide us with several answers or hypotheses, of which I’ll mention three. The Lexicalist approach claims that the morphological word is the atomic syntactic unit, the terminal node. The Incorporation approach, combined with a readjustment and spellout component, à la Halle (1989), might be taken to claim that morphological words correspond to a syntactic constituent which we might call the maximal X₀ – an X₀ which is not dominated by another X₀, thus the highest node in an adjunction structure resulting from head movement.

The differences between these two views are very subtle for the question at hand, and I won’t go into them here. On the other hand, they both are significantly different from another view, which we might associate with the works of Marantz (1988), Sproat (1988) and Sadock (1991). In this paper, I will primarily be concerned with a proposal by Sadock, regarding the treatment of the definite article in languages of the Balkan area, but many of the observations here apply equally to the other approaches. The Autolexical approach allows, under restricted circumstances, two or more syntactic nodes which don’t have to form a constituent to correspond to a single morphological word, in contrast to the other two approaches. The empirical issues this raises are most pointed in the consideration of clitics. What we will want to know is the validity, as a universal, of the following generalization:

Bound morphemes fall into two classes, often called clitics and affixes, which differ both in their distribution and their morphological or phonological properties.
As a rule of thumb, we find some bound morphemes appearing at or near the edge of the phrase they are semantically or syntactically associated with, having only minor, unexceptional interactions with their host. Others appear on the head of the phrase and can interact more strongly with it. If this generalization were universal, it would clearly support the Lexicalist or Incorporation approaches over the Autolexical account. However, proponents of the Autolexical approach point to the existence of morphemes which seem to have phrasal distribution (at or near the edge of a phrase) but do form a morphological word with their hosts. In this paper, I will examine one such case, and show that it can and should be reanalyzed in a way compatible with requiring morphological constituents to be (minimal) syntactic constituents. The analysis I propose is an extension of ideas developed in Poser (1985), Zwicky (1987), and Miller (1991).

In several languages in the Balkan area, specifically Bulgarian, Rumanian, Macedonian and Albanian, the definite article usually shows up as a suffix or enclitic to the first word of a noun phrase. The grammatical status of the article has been a point of controversy in the literature on these languages – see for instance Scatton (1980), Mayer (1988) and Elson (1976). With some exceptions, the discussion has basically been over whether the article should be considered a suffix or a clitic, with the question arising because the article apparently has the distribution of a clitic, but the phonology of a suffix. Examples (1)–(3) illustrate some basics of its distribution which suggest that it is in second position within NP. (Unless otherwise noted, all examples are taken from Bulgarian.)

1. kniga-ta  “the book”
2. xubava-ta kniga  “the nice book”
3. moja-ta xubava kniga  “my nice book”

On the other hand, it acts phonologically like a suffix. In (4), one aspect of its phonology is reviewed. There, we see that a word-level phonological rule of final-obstruent devoicing, applying in (4a), is blocked from applying before the article, as with other inflectional suffixes, as in (4b,c), while it does apply before various clitics, as in (4d,e). In this and other ways which I don’t have room to go into here, the definite article has the phonology of an inflectional suffix rather than a clitic (see, e.g., Scatton 1980).

4. a. /grad/ → [grat]; “a city”
   b. /grad+ove/ → [gradove]; “cities” (Scatton 1984, p. 128)
   c. /grad+at/ → [gradat]; “the city” (Elson 1976)
   d. /grad#e/ → [grat e]; “It is a city.” (Elson 1976)
   e. /grad#li#e/ → [grat li e]; “Is it a city?” (Elson 1976)
2. Second Position

In the rest of this paper, I will argue that, despite the data in (1)–(3), the definite article is not really a true second position (2P) clitic, and should be treated as a special type of inflection. In order to make the case against calling the article a 2P clitic, I will briefly consider the analysis of clear, uncontroversial cases of 2P so that we can compare them to the definite article.

Take the example of second position clitics in Serbo-Croatian. Serbo-Croatian sentential clitics (including auxiliary verbs, the interrogative particle, and clitic pronouns) may usually appear after the first phrase of the sentence. Alternatively, they may also appear after the first word. Furthermore, they immediately follow a complementizer which introduces a subordinate clause. These various options are illustrated in the examples in (5)–(8).

Serbo-Croatian:

(5) Visok=je čovek Petar.
tall=AUX man Petar
“Petar is a tall man.”

(6) Visok čovek=je Petar.

(7) ... pesniku koji=je napisao knjigu ove godine
... poet who=AUX wrote book this year
“... a poet who has written a book this year.”

(8) *... pesniku koji napisao=je knjigu ove godine

Various analyses have been proposed for dealing with these data, and other second position clitics. My own favorite (see Halpern 1992) works roughly as follows: second position results from an interaction of two factors. Second position clitics are subject to syntactic conditions which place them first in the appropriate domain. In the case of the Serbo-Croatian clitics, they are placed in a position which is at the beginning of a clause, except that a topicalized or focused phrase may be placed in front of them. They are also subject to a phonological condition which can result in their appearing in a surface position different from that which the syntax assigns them. This condition is a requirement that there be a phonological word to their left to which they may cliticize. This second condition will mean that if the clitic has no word preceding it in the syntax, it will be forced to move out of its syntactic position past the word which is to its right in the syntax. However, if the clitic is preceded by a word or phrase in the syntax, then its phonological condition will be satisfied without moving and it will stay put. This analysis is illustrated by the diagrams in (9)–(11). Though the notation is quite different, this analysis of 2P is basically available within the Autolexical framework.
The question to ask now is, do we want to apply this sort of analysis, or an alternative treatment, to the definite article? To a first approximation, the answer of course is "yes". That's why we started looking at 2P in the first place. The examples in (1)–(3) can be accounted for in much the same way that the Bulgarian and Serbo-Croatian clitics were treated. We can say that the definite article is a determiner which the syntax places in the same position that it puts demonstratives – at the beginning of the Noun phrase, possibly preceded by a quantifier. Unlike the demonstratives, it must attach to a word on its left and so it generally winds up on the surface after the word which follows it in the syntax, as illustrated in (12). This will account for the basic examples where the clitic appears after the first word. (13) shows an example of the situation with demonstratives according to this account. This analysis, notation aside, is a slight simplification of that which Sadock (1991) made for the Macedonian definite article.
3. Definite article positioning: fine points

However, when we look more closely, we find that there are certain respects in which this analysis is problematic. For one thing, uncontroversial cases of second position clitics seem restricted to appear in one of the positions available to the Serbo-Croatian clitics. That is, they end up after the first phonological word (or some other prosodic unit) of their domain, or after the leftmost syntactic daughter of that domain, but not somewhere in the middle of this first daughter.

In the examples of the definite article discussed so far, either of these descriptions is applicable. However, when we examine various complicated cases, problems come up. When a noun phrase begins with an adjective or deverbal adjective modified by a prepositional phrase, the article comes immediately after the adjective; this shows that it can’t be right to say that the article can appear after the first phrase, though it is consistent with saying that it appears after the first word. When we look at other examples, it is clear that the article isn’t always after the first word either, and in fact sometimes must appear in positions which are neither after the first word nor after the first phrase. This can be seen in the examples in (14)–(17). To emphasize the significance of this, no clear second position clitic which I know of has this distribution, giving us reason to look more carefully at the definite article.

(14) silno razprostranena-ta upotreba
strongly widespread-DEF use
“the very widespread use”
(Spostavitelno ezikoznanje #2, 1990 p. 14)

(15) vse oshter sporni-te kriterii
always yet controversial-DEF criteria
“the still controversial criteria” (p. 110)

(16) sostojala-ta se prez vreme na konferencija-ta okrąglə masa
occurred-DEF REFL during time of conference-DEF round table
“the round table which took place during the conference” (p. 85.)

(17) pochti nerazrabotena-ta u nas problematika
almost undeveloped-DEF among us problem.area
“the problem area which is almost undeveloped among us” (p. 90)

Another problem for this sort of analysis has to do with the occurrence of multiple copies of the article within a single NP or with cooccurrence of the article and a demonstrative. The second position analysis given above entails that only a single copy of the article should appear per NP, and that the article and demonstratives should be in complementary distribution: the presence of one should block the presence of the other. Both of these predictions are reasonably accurate for Bulgarian, but there are a couple of circumstances in which they are problematic, and a look at the situation in Macedonian and Rumanian makes it clear that they are wrong in the general case.

In Bulgarian, in most cases, neither a demonstrative nor other words in an NP with a demonstrative may host the definite article, as shown in (18a–c). However, in a noun phrase which has a quantifier before a demonstrative, it is possible to
also have the definite article on the quantifier, as in (18d,e). The second position analysis makes this unexpected.

(18)  
  a. tezi profesori  \quad \text{"these professors"}
  b. *tezi-te profesori  \quad \text{"these-the professors"}
  c. *tezi profesori-te  \quad \text{"these professors-the"}
  d. vsički tezi profesori  \quad \text{"all these professors"}
  e. vsički-te tezi profesori  \quad \text{"all-the these professors"}

Data regarding the definite article in Macedonian, which is much like the article in Bulgarian, supports this argument. At least in certain dialects of Macedonian, it is possible to have the definite article cooccur with the demonstrative even if nothing precedes demonstrative. (19) give some basic Macedonian examples, while (20) illustrates the point at hand.

Macedonian: (Lunt 1952, Sadock 1991)

(19)  
  a. čovek-ot  \quad \text{"the man"}
  b. malijot čovek  \quad \text{"the little man"}
  c. dobrijot mal čovek  \quad \text{"the good little man"}

(20)  
  Ona malki-te Grci shto gi imashe...
  Those few-DEF "Greeks" that there were...
  "Those few Greeks that there were"

As for multiple copies of the definite article, once again the clitic analysis is generally right, but there are occasional counterexamples. Mayer (1988) discusses the situation in Bulgarian when a noun phrase begins with a set of coordinated adjectives. Most of the time, only the first adjective will receive the definite article, but in certain cases it may appear on each conjunct. Mayer argues that repeated articles result from full noun phrases having been coordinated and with part of the first one getting deleted. An elliptical analysis of this sort seems reasonable given the meaning of the examples involved, so if we only look at Bulgarian, this may not be problematic. Examples, from Mayer (1988:73), are given in (21).

(21)  
  a. xubava-ta i plodorodna zemja  
      \quad \text{"the beautiful and fertile land"}
  b. balgarski-te i savetski-te studenti  
      \quad \text{"the Bulgarian and the Soviet students"}

However, looking at some data from Rumanian, it becomes clear that this does not fully solve the problem. In Rumanian once again, definiteness can be expressed by a suffix or enclitic roughly in second position, as in (22)–(25). Contrary to what we would expect if it were truly a clitic, when we look at NP’s involving coordination, we find that the definite article must appear on each conjunct. That is, duplication of the article is obligatory if a noun phrase begins with a coordinate structure, even when the meaning is not consistent with an elliptical analysis. Moreover, the article will sometimes appear more than once even in the absence of coordination. Examples illustrating these points are given in (26) and (27).
Rumanian: (Lombard 1974, Grosu 1988)

(22) poé-tu-l märe
    poet-DEF great
    "the great poet"

(23) märe-le poét
great-DEF poet
    "the great poet"

(24) senzaţional-a nouă idee
    sensational-DEF new idea
    "the sensational new idea"

(25) foarte frumosul cal
    very beautiful-DEF horse
    "the very beautiful horse"

(26) mîndra (și) frumoasa femeie
    proud-DEF (and) beautiful-DEF woman
    "the proud and beautiful woman"

(27) frumosul, noul vas
    beautiful-DEF new-DEF vase
    "the beautiful new vase"

Similar cases of definite articles appearing more than once in a single noun phrase occur in Greek (C. Condoravdi, p.c.) as well as the unrelated case of Amharic (Halefom 1990). It appears then to be a general property of definite articles of this type that they can appear more than once in a single NP, unlike what is observed of true second position clitics.

4. Analysis

In the light of these arguments against treating the article as a second position clitic, let us consider an alternative analysis. In line with the phonological and morphological evidence for treating the article as a suffix, suppose we claim that it is actually an inflectional affix. Of course, the immediate problem is that as an inflection pertaining to noun phrases, we expect to find it on the head of the NP. I suggest that this is simply too limited a view of inflection, and that the resolution to the paradox posed by the article is not to allow clitics to have phonology of affixes, but to allow affixes to appear in other locations. My proposal is based in part on ideas in Halpern and Miller (in prep) and Svenonius (1991).

First, we have to distinguish two things: the semantic or syntactic property of being definite (represented as the specification “def” in the following diagrams), which is a property of the entire noun phrase, and the morphological property of having the definite suffix (represented as the specification “def-morph”), which is primarily, though we will see not exclusively, a property of lexical items. The two are obviously intimately related, and we can treat the syntactic property as responsible for triggering the morphological property, as in the implicational rule in (28), which I will refer to as “Rule 1”. This rule says that if a noun phrase is syntactically definite then its leftmost daughter is morphologically definite.3
(28) Balkan definiteness, Rule 1

\[
\begin{array}{c}
\text{NP[def]} \\
\text{X(P)} \\
\text{(...)} \\
\rightarrow \\
\text{NP[def]} \\
\text{X(P)[def-morph]} \\
\text{(...)}
\end{array}
\]

For the sake of comparison, consider the corresponding rule in (29) which we might write for the English plural. It says that a syntactically plural noun phrase should have plural morphology on its head. The difference between the two rules is that the first requires the leftmost daughter bear certain morphology while the second imposes that specification on the head.

(29) English plural rule

\[
\begin{array}{c}
\text{NP[pl]} \\
\text{(...)} \\
\text{N} \\
\text{(...)} \\
\rightarrow \\
\text{NP[pl]} \\
\text{(...)} \\
\text{N[pl-morph]} \\
\text{(...)}
\end{array}
\]

Rule 1 accounts for the distribution of the definite article in the simple case where the first member of the noun phrase is a single word by requiring that word to bear definite morphology. Once this is determined, the morphology treats this specification the way it treats any specification about inflection: it adds the appropriate affix to a stem or word, in this case an allomorph of the definite article. The result is illustrated in (30)–(32).

(30) \[
\begin{array}{c}
\text{NP[def]} \\
\text{N[def-morph]} \\
\text{kniga-ta}
\end{array}
\]

(31) \[
\begin{array}{c}
\text{NP[def]} \\
\text{A[def-morph]} \\
\text{xubava-ta} \\
\text{N} \\
\text{kniga}
\end{array}
\]

(32) \[
\begin{array}{c}
\text{NP[def]} \\
\text{A[def-morph]} \\
\text{moja-ta} \\
\text{xubava} \\
\text{N} \\
\text{kniga}
\end{array}
\]

On the other hand, when the first modifier of the noun phrase is a phrase itself, Rule 1 will result in the modifier phrase as a whole bearing definite morphology. Rule 2, in (33) indicates how this is to be interpreted: if a phrase is definite marked (that is, is specified "def-morph"), then its head must be. This seems to be the right result, in light of cases like examples (14)–(17) above. In all of these examples, the article appears on the adjective which is the head of the first daughter of the NP. The analysis given to (14) is shown in (34).
Balkan definiteness, Rule 2

(33) \[
\begin{array}{c}
\text{XP[def-morph]} \\
(\ldots) \quad X \quad \text{(\ldots)}
\end{array}
\rightarrow
\begin{array}{c}
\text{XP[def-morph]} \\
(\ldots) \quad X[\text{def-morph}] \quad \text{(\ldots)}
\end{array}
\]

(34) \[
\begin{array}{c}
\text{NP[def]} \\
\text{AP[def-morph]} \quad \text{N} \\
\text{ADV} \\
\text{silno} \quad \text{razprostranena-ta} \quad \text{upotreba} \quad (= (14))
\end{array}
\]

This is the essence of my proposal regarding the Balkan article: Rules 1 and 2 determine which word in a noun phrase must be definite-marked and inflectional morphology determines the forms of words which are so specified.

There are three extensions which I will now turn to. The first has to do with the demonstratives. As noted above, demonstratives and the definite article don’t generally cooccur in Bulgarian, but they can do so in a certain environment, namely when the demonstrative is preceded by a quantifier. This is evocative of another class of words which don’t take the definite article themselves, but may cooccur with it if preceded by a modifier, namely kinship terms for close family members. See (35)-(37).

(35) žena ambig: “a woman/wife” or “the (my) wife”
(36) *žena-ta intended: “the (my) wife”
(37) moja-ta žena “my wife”

We can explain this as follows. One consequence of treating the definite article as an inflectional suffix rather than a clitic is that it will be possible to have exceptionally inflected forms. We can say that while the definite form of most words is formed by adding the definite suffix, some words are their own definite form, just as in English the noun “sheep” is its own plural or the verb “hit” is its own past tense. We can say then that the kinship terms and the demonstratives are their own definite forms. When they are initial in a definite NP, rule 1 will force them to appear in their definite form, but this will not result in any visible change. However, when they are not initial in an NP, the definiteness marking will pass onto another form, as in (38)–(41).

(38) \[
\begin{array}{c}
\text{NP[def]} \\
\text{N[def-morph]} \\
\quad \text{žena}
\end{array}
\]
(39) \[
\begin{array}{c}
\text{NP[def]} \\
\text{A[def-morph]} \\
\quad \text{inteligentna-ta}
\end{array}
\]
(40) \[
\begin{array}{c}
\text{NP[def]} \\
\text{N} \\
\quad \text{žena}
\end{array}
\]
Support for the claim that demonstratives as well as kinship terms can be their own definite forms comes from the behavior of a set of clitic possessive pronouns. These clitic pronouns generally can only follow a word with the definite article attached to it, as in (42) and ((43). However, (44) and (45) show that they may, apparently exceptionally, follow the kinship terms without an article, and at least marginally may also follow demonstratives.

\[
(42) \quad *\text{in\'elligentna mi \'zena} \quad \text{"my intelligent wife"}
\]
\[
(43) \quad \text{in\'elligentna-ta mi \'zena} \quad \text{"my intelligent wife"}
\]
\[
(44) \quad \text{\'zena mi} \quad \text{"my wife"}
\]
\[
(45) \quad ? \text{ tazi mi k\'o\'sta} \quad \text{"that house of mine"}
\]

The second issue has to do with the behavior of the article and coordinated adjectives. As I said, the usual result in Bulgarian is for the article to appear only on the first conjunct, but in Rumanian it must always appear on all conjuncts. It seems that we have to say that rule 2 is interpreted differently in the two languages for coordinate structures, so that it puts the definite morphology specification on only the left conjunct in Bulgarian but on all conjuncts in Rumanian. Of the two, the pattern in Bulgarian is probably the more marked situation, but this interpretation is supported by the existence of other cases cross-linguistically where inflections appear on only one conjunct in a coordinate structure, leaving the others in an uninflected form – see Payne (1985).

One final point is that there are certain adjectives, primarily of foreign origin, which can't serve as hosts for the definite article; furthermore, when they are the first modifier in an NP, the article can't appear on another word. This is illustrated in (46) and (47). (48) shows that these adjectives may be used in definite NP's, as long as they would not be expected to host the definite article.

\[
(46) \quad *\text{serbez-ta} \quad \text{\'zena}
\]
\[
\text{quarrelsome-DEF} \quad \text{woman}
\]

intended: "the quarrelsome woman"
(47) *serbez žena-ta
quarrelsome woman-DEF
intended: "the quarrelsome woman"

(48) tazi serbez žena
that quarrelsome woman
"that quarrelsome woman"

These adjectives turn out to be generally defective with respect to nominal inflection, showing either reduced inflection for number and gender or none at all, as shown in (49) and (50).

(49) serbez čovek
quarrelsome man
"a quarrelsome man"

(50) serbez žena (cp. *serbeza žena)
quarrelsome woman
"a quarrelsome woman"

All that we need to say to account for this is that these adjectives are syntactically unexceptional but that they lack a form inflected for the definite article, preventing their insertion under a node specified for "def-marking". Such gaps in an inflectional paradigm are unremarkable, but it would be an odd result for a clitic.

5. Conclusion

To summarize, the definite article in several Balkan languages acts like an inflectional suffix phonologically and morphologically. As for its distribution, three arguments were made against the obvious characterization of the article as a second position clitic. First, in NP’s with complex modifiers it turns out that the article appears, as a rule, neither after the first word nor the first branch of the NP. Second, it is possible for the article to cooccur with demonstratives in certain circumstances. Third, multiple copies of the article can appear in a single NP. These facts may be better explained if the article is treated as an inflection whose distribution is determined by the two rules given in (28) and (33). This account also gives us a way of treating the exceptional properties of kinship terms and demonstratives with respect to the clitic possessive pronouns, and accounts for the behavior of certain morphologically defective adjectives. I conclude that the Balkan definite article is an argument in favor of allowing inflectional suffixation away from the head of a phrase, and indirectly an argument against allowing clitics to interact strongly, in phonological terms, with their hosts.

Notes

1 I would like to thank Wayles Browne, Donka Farkas, Victor Friedman, Sharon Inkelas, Philip Miller, Bill Poser, Ivan Sag, Ernie Scatton, Maxim Stamenov, Peter Svenonius, Draga
Zec and Arnold Zwicky for discussion and aid in various ways. All errors and lunacies are mine.

2 Note that this is not the same question as that of the relationship between the phonological or prosodic word and the syntax.

3 I am calling this a 'rule', and writing it as such, but ultimately it is better treated as a feature instantiation principle along the lines of the Head Feature Convention of GKPS (1985).

References
Svenonius, Peter. 1991. The Definite Suffix in Norwegian as a Phrasal Affix. ms. UCSC.