Inward-sensitive contextual allomorphy and its conditioning factors

Boris Harizanov & Vera Gribanova

University of California, Santa Cruz & Stanford University

1. Introduction

In a representation like (1), the form of the morpheme M₂ can be determined by another morpheme that is either closer to the root than M₂ (e.g. M₁), or farther away from the root than M₂ (e.g. M₃)—inward vs. outward allomorphic sensitivity.

(1) [ROOT]–[M₁]–[M₂]–[M₃]

In this paper we use a case study from the paradigm of Bulgarian definiteness marking to investigate whether the direction of allomorphic sensitivity is correlated with the type of information (phonological, morphosyntactic) that conditions this allomorphy. We evaluate this question against a theoretical backdrop of three assumptions commonly entertained within Distributed Morphology (DM; e.g. Bobaljik 2000): (i) Separation: morphology interprets syntax; i.e. “late” insertion; (ii) Cyclicity: the insertion of phonological material proceeds root-outwards; and (iii) Rewriting: as morphosyntactic features are expressed by phonological material, these features are used up and no longer part of the representation. Assumptions (i) and (ii) are standard within DM (Halle and Marantz 1993), and (iii) is easily accommodated within such a theory (but compare e.g. Halle 1990 and Bobaljik 2000 with Halle and Marantz 1993). Taken together, these assumptions yield two predictions about allomorphic behavior: outward sensitive allomorphy can only be conditioned by morphosyntactic features, while inward-sensitive allomorphy can only be conditioned by phonological features.

Our claim in this paper is that both morphosyntactic and phonological features are relevant for inward-sensitive allomorphy of the Bulgarian definiteness marker (§2; that is,

*The authors would like to thank, for their generous advice and feedback on this project: Karlos Arregi, Ryan Bennett, Lev Blumenfeld, Sandy Chung, Amy Rose Deal, David Embick, Steven Franks, Jorge Hankamer, Paul Kiparsky, Ruth Kramer, Ora Matushansky, Jim McCloskey, Jason Merchant, Andrew Nevins, Mark Norris, Peter Svenonius, Matt Tucker, Draga Zec, the audience at the Workshop on Locality and Directionality at the Morphosyntax-Phonology Interface at Stanford University, and the participants of UCSC’s Morphology Reading Group and Crosslinguistic Investigations in Syntax-Phonology. All errors are the authors’ responsibility.
that assumptions (i), (ii) and (iii) cannot all be maintained—contra Bobaljik 2000). In §3, we explore and compare two versions of lexical insertion which are consistent with (i) and (ii), but reject the strongest version of (iii). Finally, in §4 we defuse two potential objections to our interpretation of the data.

2. **Definiteness marking in Bulgarian**

The Bulgarian definiteness marker (DEF) follows the left-most nominal head within a definite nominal phrase (2).¹

(2) a. staro-
old-DEF
to
dârvo
‘the old tree’
b. tri-
three-DEF
 te
novi
ew books
knigi
‘the three new books’
c. tvârde
excessively heavy-DEF
težka-
heavy-
 ta
masa
‘the excessively heavy table’
d. prohladna-
cool-
 ta
i
and
eveča-
fresh
evning
sveža
večer
‘the cool and fresh evening’

DEF is a phonological suffix in the sense that it forms a prosodic word with the nominal head to its left. For example, just like inflectional suffixes (3-b) but unlike auxiliary clitics (3-c), it bleeds word-final devoicing (3-d):

(3) a. /bratovčed/ → [bratofčet] ‘cousin’
b. /bratovčed + iplural/ → [bratofčedi] ‘cousins’
c. /bratovčed + e
kopula/ → [bratofčet#e] ‘it’s a cousin’
d. /bratovčed + aDEF/ → [bratofčeda] ‘the cousin’

DEF surfaces in one of five forms: -a, -to, -te, -ta, or a stress-attracting -tá. These changes in surface form instantiate inward-sensitive allomorphy, as DEF is uniformly the most peripheral suffix on a nominal head:

(4) [ROOT]—[DERIVATION]—[NUMBER,GENDER]—[DEF]

As elaborated in the next section, the form of DEF depends on both morphosyntactic and phonological properties of the stem to which it attaches (e.g. Scatton 1984, Franks 2001).

2.1 **Morphosyntactically sensitive allomorphy**

The form of DEF is determined in part by the gender and number features of its morphological host. For example, the majority of masculine singular nouns take the -a allomorph (see

¹By “nominal” here we mean any head that exhibits number/gender concord within a nominal phrase.
Inward-sensitive contextual allomorphy

(12) and (13) in §2.2 for the exceptions). So do all masculine adjectives and other nominal modifiers: ²

(5) **Singular masculine**

a. mąż ‘man’ — mąża ‘the man’
   b. učitel ‘teacher’ — učitelja
   c. čajnik ‘teapot’ — čajnika
   d. kraj ‘end’ — kraja
   e. visok ‘tall’ — visokija ‘the tall’
   f. dobår ‘good’ — dobrijà
   g. tutkav ‘slow’ — tutkavija
   h. ovči ‘sheep (attr)’ — ovčija

Most feminine singular nouns and all feminine singular adjectives end in -a and take the -ta allomorph:

(6) **Singular feminine (vowel final)**

a. žena ‘woman’ — ženata ‘the woman’
   b. učitelka ‘teacher’ — učitelkata
   c. voda ‘water’ — vodata
   d. ideja ‘idea’ — idejata
   e. visoka ‘tall’ — visokata ‘the tall’
   f. dobra ‘good’ — dobrata
   g. tutkava ‘slow’ — tutkavata
   h. ovča ‘sheep (attr)’ — ovčata

However, some feminine singular nouns end in a consonant and instead take the stress-attracting -tá allomorph: ³

(7) **Singular feminine (consonant final)**

a. pesen ‘song’ — pesentá
   b. mladost ‘youth’ — mladosttá
   c. hubost ‘beauty’ — hubosttá
   d. cev ‘barrel (of a gun)’ — cevtá

All neuter singular nouns and adjectives take the -to allomorph:

(8) **Singular neuter**

a. oko ‘eye’ — okoto ‘the eye’
   b. dete ‘child’ — deteto
   c. bižu ‘jewel’ — bižuto
   d. žuri ‘jury’ — žurito
   e. visoko ‘tall’ — visokoto ‘the tall’
   f. dobro ‘good’ — dobrotó
   g. tutkavo ‘slow’ — tutkavoto
   h. ovče ‘sheep (attr)’ — ovčeto

Finally, in the plural—where gender distinctions are neutralized—DEF is realized as -te with both nouns and adjectives, in many, but not all cases (see (14) in §2.2 on DEF’s alternative realization in the plural):

(9) **Plural**

a. maže ‘men’ — mažete ‘the men’
   b. učitel ‘teachers (M)’ — učitelte
   c. učitelki ‘teachers (F)’ — učitelkite
   d. oči ‘eye’ — očite
   e. visoki ‘tall’ — visokite ‘the tall’
   f. dobri ‘good’ — dobrite
   g. tutkavi ‘slow’ — tutkavite
   h. ovči ‘sheep (attr)’ — ovčite

²In the standard language, the masculine “adjectival stem extension” -i- appears with adjectives that do not contain it already.

³In §3 we treat this behavior as an instance of (partially) phonologically conditioned allomorphy; on the role of phonology in allomorphy, see §2.2.
Importantly, in some cases, when DEF attaches to the head noun in a nominal phrase, its form is attributable to the gender feature, and not the phonological form, of the noun. As (10) and (11) demonstrate, homonyms may take distinct allomorphs of DEF. Thus, phonological information is not a sufficient conditioning factor; reference to gender features is crucial.

(10) a. gaz (M) ‘gas (state of matter)’ — gaza ‘the gas’
   b. gaz (F) ‘gas (fuel)’ — gaza.ta ‘the gas’
(11) a. med (M) ‘honey’ — meda ‘the honey’
   b. med (F) ‘copper’ — medta ‘the copper’

2.2 Phonologically sensitive allomorphy

Gender and number features alone are not sufficient for determining the form of DEF either; the phonological shape of DEF’s host is also crucial. For example, a small set of vowel-final masculine singular nouns which end in -a or -o, take the -ta and -to allomorphs of DEF, respectively (not -a):

(12) *Singular masculine, final -a* (cf. (5))
   a. bašta ‘father’ — bašata ‘the father’
   b. sâdija ‘judge’ — sâdíjata ‘the judge’
   c. lovđija ‘hunter’ — lovđijata ‘the hunter’
   d. bojadžija ‘painter’ — bojadžijata ‘the painter’
(13) *Singular masculine, final -o* (cf. (5))
   a. tatko ‘dad’ — tatko.to ‘the dad’
   b. djado ‘grandfather’ — djado.to ‘the grandfather’
   c. čičo ‘uncle’ — čičoto ‘the uncle’
   d. vujčo ‘uncle’ — vujčoto ‘the uncle’

Additionally, some pluralizing suffixes end in -a; these are followed by the -ta allomorph, instead of -te:

(14) *Plural, final -a* (cf. (9))
   a. bratja ‘brothers’ — bratjata ‘the brothers’
   b. pâtišta ‘roads’ — pâtišata ‘the roads’
   c. moreta ‘seas’ — moretata ‘the seas’
   d. taksita ‘taxi cabs’ — taksitata ‘the taxi cabs’

Finally, distinct plural forms of the same noun occur with different allomorphs of DEF:

(15) a. kolenata, kolene ‘knees’
   b. kolenate, kolenete ‘the knees’
(16) a. ramena, ramene ‘shoulders’
   b. ramenata, ramene ‘the shoulders’
Inward-sensitive contextual allomorphy

Based on this evidence, we conclude that number and gender features do not uniquely determine allomorph selection.

3. Analysis

The distribution of the five allomorphs of DEF can be informally described by the following (ordered) if... then statements, which make reference to both morphosyntactic and phonological information:

(17) a. If the stem ends in a consonant and
   (i) if the stem is singular masculine, then DEF → -a
      brat-a ‘the brother’, čaj-a ‘the tea’, moliv-a ‘the pencil’
   (ii) if the stem is singular feminine, then DEF → -tá
      kráv-tá ‘the blood’, doblest-tá ‘the valor’, prolet-tá ‘the spring’
   b. Otherwise (i.e. the stem ends in a vowel),
      (i) if the stem ends in -a, then DEF → -ta
         bašt-a-ta ‘the father (M)’, žena-ta ‘woman (F)’, kraka-ta ‘the feet (PL)’
      (ii) if the stem is plural, then DEF → -te
         máže-te ‘the men’, ženi-te ‘the women’, sto-te ‘the hundred’
      (iii) otherwise, DEF → -to
         more-to ‘the sea (N)’, taksi-to ‘the taxi (N)’, tatko-to ‘the dad (M)’

Next, we explore two distinct implementations of this algorithm for determining the form of DEF in the general framework of DM.

3.1 One solution

Recall that, according to one view of lexical insertion, as morphosyntactic features are expressed by phonological material, these features are used up and no longer part of the representation—the Rewriting assumption from §1 (Halle 1990:156, Noyer 1992:23, Bobaljik 2000:16). Any system that incorporates this assumption undergenerates with respect to the allomorphy exhibited by DEF, which is inwardly sensitive to morphosyntactic information. Therefore, one solution is to reject Rewriting (Halle and Marantz 1993, Embick 2010:39), so that lexical insertion has simultaneous access to both morphosyntactic and phonological context. Under this view, the following would be legitimate Vocabulary Items:

(18) Vocabulary Items
    a. [DEF] ↔ -a / -C#, [SG, MASC] __
    b. [DEF] ↔ -tá / -C#, [SG, FEM] __
    c. [DEF] ↔ -ta / -a# __
    d. [DEF] ↔ -te / [PL] __
    e. [DEF] ↔ -to ___

---

4Here, “#” signals the right edge of the stem to which DEF attaches and “C” stands for “consonant”.
In both (19-a) and (19-b), for example, D[DEF] is realized by the allomorph in (18-c) (i.e. -ta because the final segment of its host is -a):\(^5\)

\[
(19) \quad \begin{align*}
&\quad \sqrt{ba\check{\text{s}}ta}\text{-}-ta \, \text{‘the father’} \\
&\quad D \\
&\quad \quad n/\text{Num} \\
&\quad \quad D \\
&\quad \quad \quad ta \\
&\quad \quad n/\text{Num} \\
&\quad \quad [\text{SG, MASC}] \\
&\quad ba\check{\text{s}}ta \quad \emptyset \\
&\quad n/\text{Num} \\
&\quad D \\
&\quad \quad ta \\
&\quad n/\text{Num} \\
&\quad [\text{PL}] \\
&\quad kraka \quad a \\
\end{align*}
\]

In DM, the most highly specified Vocabulary Item whose identifying features are a subset of the features of the terminal node is inserted at this terminal node (Halle and Marantz 1993). When competing Vocabulary Items match the same number of features of the terminal node (just one in this case, [DEF]), the context for insertion of a Vocabulary Item becomes relevant. Consequently, the -a allomorph (18-a) is not inserted in (19-a), since the stem does not end in a consonant, even though it does bear singular masculine features and the contextual specification of (18-a) is partially matched.

A difficulty arises with respect to the competing Vocabulary Items -ta (18-c) and -te (18-d) in (19-b). They match the same number of features of the terminal node (just one, [DEF]) and their context specifications are both matched. Normally, the principle governing choice between competing Vocabulary Items in DM (Subset Principle, Halle and Marantz 1993) always chooses the one with the most specific context of insertion. However, how is competition regulated between Vocabulary Items with morphosyntactic context and Vocabulary Items with phonological context when both types of contexts are actually matched? A possibility explored in Harizanov and Gribanova 2011, and taken up here, is that the Subset Principle be further articulated by specifying that phonological context (a-final stem) is more specific for the purposes of allomorph selection than morphosyntactic context (plural stem). This will ensure that -ta (18-c) is taken at lexical insertion to be more specific than -te (18-d) in (19-b).

### 3.2 Another solution

Bye and Svenonius (2012) develop an alternative lexical insertion procedure consisting of two steps: (i) L-match, which associates possible allomorphs to a morpheme M according to M’s features and morphosyntactic context; and (ii) Insert, which selects a unique allomorph on purely phonological grounds. For D[DEF] in the context of ba\check{s}ta ‘father’, this system chooses the set of allomorphs \{-a, -ta, -to\} as possible matches, leaving the phonological component to decide on the actual exponent -ta (based on the phonological shape of the stem):

---

\(^5\)These trees represent morphophonological structure and encode linear order; they are the output of whatever morphosyntactic mechanisms are responsible for the placement of the (arguably) mobile D[DEF].
Inward-sensitive contextual allomorphy

(20) a. Step 1: L-match

\[
\begin{align*}
D & \\
| \quad n/\text{Num} & \quad D \\
| \quad N & \quad n/\text{Num} \\
\text{bašt\a} & \quad \emptyset
\end{align*}
\]

b. Step 2: Insert

\[
\begin{align*}
D & \\
| \quad n/\text{Num} & \quad D \\
| \quad {a, \text{ta, to}} & \quad \{\text{DEF}\} \\
| \quad N & \quad n/\text{Num} \\
\text{bašt\a} & \quad \emptyset
\end{align*}
\]

What this implementation has in common with the one described in §3.1 is that it allows lexical insertion to reference inward morphosyntactic information even though this information has already been associated with phonological exponent.

4. Two potential objections

Two objections to our interpretation of the Bulgarian data can be raised which, if accurate, might suggest—contrary to what we have claimed above—that inward-sensitive allomorphy can be conditioned only by morphosyntactically irrelevant (morphophonological) features that are provided at lexical insertion (Vocabulary Insertion in DM; henceforth, VI). A proponent of such a view might argue, first, that number and gender in Bulgarian should be considered morphophonological features that are morphosyntactically irrelevant and inserted only at VI. A second argument might be that DEF allomorphy is not inward-sensitive at all—e.g. DEF bears the relevant number and gender features itself. However, neither of these conjectures receives empirical support in Bulgarian.

4.1 The status of number and gender

We consider three pieces of evidence that number and gender in Bulgarian nominal phrases are morphosyntactic features, available before VI. First, Bulgarian number and gender are highly productive and fairly regular in their exponent; in DM, this means that they should be present in the syntax, and independent from the root/stem. Plural nouns in Bulgarian can generally be decomposed into a root/stem and a discrete (usually) non-null plural suffix:

(21) a. māz ‘man’ – māže ‘men’
    b. koljano ‘knee’ – kolene
    c. učitel ‘teacher’ – učitel\i
    d. žena ‘woman’ – ženi ‘women’
    e. dvor ‘yard’ – dvorove ‘yards’
    f. grad ‘city’ – gradove ‘cities’

Parallel to number, gender usually receives consistent morphophonological expression separate from the root, as reflected in subject-predicate agreement and concord:

\textit{It has been suggested that idiosyncratic properties of particular roots, such as declension class or phonological features, are not syntactically relevant and should, therefore, be able to condition inward-sensitive allomorphy. However, this claim has been challenged on the basis of the existence of idiosyncratic properties of roots that, nonetheless, participate in syntactic processes (e.g. Embick 2000). In addition, there is abundant evidence that (at least) number and interpretable (i.e. natural) gender features in Bulgarian are not idiosyncratic properties of roots (see below).}
In addition, while inanimate roots are idiosyncratically licensed in the context of particular genders, (23) shows that many animates occur with distinct gender suffixes, reflecting the referent’s natural gender. This fact further corroborates the syntactic independence of (at least interpretable natural) gender features.

(23) a. pevec ‘singer (M)’ — pevica ‘singer (F)’
   b. bălgarîn ‘bulgarian (M)’ — bălgarka ‘bulgarian (F)’
   c. aktjor ‘actor (M)’ — aktrisa ‘actress (F)’

A second argument for the independent morphosyntactic behavior of number and gender is that they both participate in nominal concord (24) and predicate-argument agreement (25); both processes are typically assumed to occur before VI, even if they are assumed to be post-syntactic, as in Bobaljik 2008. From this we conclude that number and gender features must be available before VI, for agreement and concord to manipulate.

(24) a. nov-a knig-a
    new-SG.F book-SG.F
   b. nov-i knig-i
    new-PL book-PL

(25) a. tazi knig-a e nov-a
    this.SG.F book-SG.F is new-SG.F
   b. tezi knig-i sa nov-i
    these book-PL are new-PL

A final argument for the morphosyntactic relevance of Bulgarian number and gender is that both have semantic (interpretive) effects, which implies that they are not purely morphophonological. Consider, for example, gender mismatches between a binder and the pronoun it binds:

(26) Vseki členj na otbora ni pomoli da i platim.
    every.SG.M member.SG.M of the.team us asked to 3.SG.F pay
    ‘Every member of the team asked us to pay her.’

Above, the binder is masculine (as reflected in concord and predicate agreement) while the bound pronoun is feminine. The bound pronoun reading necessarily involves an all-female team, and the only source of this interpretation is the feminine gender feature on the pronoun (gender on the pronoun cannot be the result of agreement because člen ‘member’ triggers masculine agreement). We conclude that since the gender feature has an interpretive effect, it should be available prior to VI.

To summarize, the productivity and regularity of number and gender exponence, their role in nominal concord and argument-predicate agreement, and their interpretive consequences all suggest that these features are relevant to the morphosyntax and therefore should be present in the computation prior to VI.
4.2 The Bulgarian definiteness marker and nominal concord

Here, we demonstrate that DEF allomorphy is necessarily inward-sensitive to the number and gender features of DEF’s host; i.e. DEF does not bear these features itself. Assume for a moment the contrary, that DEF actually spells out a D morpheme that contains the features [DEF, NUMBER, GENDER]:

(27) Alternative analysis (cf. (19-a) and (19-b))

\[
\text{D} \quad \text{n/Num} \quad \text{D} \quad \text{[DEF, NUM, GEN]}
\]

\[
\text{N} \quad \text{n/Num} \quad \text{[NUM, GEN]}
\]

Perhaps, DEF acquires these features via concord—of the type that transmits features to adjectives—before it is suffixed in the morphophonology:

(28) Feature transmission to A (adjectival concord)

\[
a. \quad \text{AP} \quad \text{NP} \quad \text{[NUM, GEN]} \\
\quad \ldots \text{A} \ldots
\]

\[
b. \quad \text{AP} \quad \text{NP} \quad \text{[NUM, GEN]} \\
\quad \ldots \text{A} \ldots \quad \text{[NUM, GEN]}
\]

(29) Feature transmission to D (cf. adjectival concord)

\[
a. \quad \text{DP} \quad \text{D} \quad \text{NP} \quad \text{[DEF, NUM, GEN]}
\]

\[
b. \quad \text{DP} \quad \text{D} \quad \text{NP} \quad \text{[DEF, NUM, GEN]}
\]

However, this approach faces both conceptual and empirical difficulties. First, the necessary relation between D[DEF] and the source(s) of number and gender feature values does not seem to be motivated (see Bonet and Harbour’s (2012) discussion of Kiowa). Specifically, D[DEF] differs from the elements that normally undergo concord in Bulgarian (i.e. nominal modifiers) in both phrase structural and morphophonological status: while D is a head in the extended nominal projection that is spelled out by a phonologically dependent unit, elements that undergo concord are either adjuncts or specifiers.

The second difficulty for this approach is empirical: the process that determines the form of DEF in Bulgarian is more local than concord. To demonstrate this, we consider configurations involving coordinated singular adjectives that cooccur with a plural head noun:
In this case, D[DEF] attaches to the left-most adjective, as expected—see (2-d). Crucially, it surfaces as -a, -to, or -ta, not -te:?

If number/gender concord affected D[DEF] the same way it affects adjectives, as suggested in (29), one would expect D to be plural (i.e. -te in the example above). However, as shown in (31), it is not. Therefore, (31) cannot be associated with the following representation, where plural features are transmitted to D via concord:

On the other hand, it is clear that in such a configuration, concord does transmit plural features to a higher adjective (*prijatelsk-ij-a ‘friendly’ in the example below), in accordance with (28):

These examples involve a single DP and, thus, a single D head. If, instead, there are two D[DEF]’s—bălgarsk-o-to i grăck-o-to pravitelstv-a ‘the Bulgarian and the Greek governments’—it is DPs that are being coordinated and not As.
Inward-sensitive contextual allomorphy

Example (33-a) is an instance of genuine number/gender concord, associated with the underlying representation in (34). Therefore, we conclude that number/gender concord cannot be responsible for the form of DEF.

(34)

In this paper, we have presented evidence from the Bulgarian definiteness marking paradigm that inward-sensitive allomorphy must be able to make reference to morphosyntactic features (number and gender) in addition to the phonological shape of the conditioning stem. To model such data in a “late” insertion theory of morphology, lexical insertion (and, thus, allomorph selection) must be able to reference both types of information. We have presented two approaches that are able to incorporate this observation; although they are distinct in some of their assumptions, what they have in common is that they allow the form of a morpheme M to be conditioned by both the morphosyntactic and phonological properties of material that is closer to the root than M. It remains to be seen whether they can be teased apart on other grounds.

5. Conclusion

References


Bye, Patrick, and Peter Svenonius. 2012. Non-concatenative morphology as epiphe-


Boris Harizanov
Department of Linguistics
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064
http://people.ucsc.edu/~bharizan/

bharizan@ucsc.edu

Vera Gribanova
Department of Linguistics
Stanford University
Margaret Jacks Hall, Building 460
Stanford, CA 94305-2150
http://www.stanford.edu/~gribanov/

gribanov@stanford.edu