# UC Santa Cruz

## Morphology at Santa Cruz: Papers in Honor of Jorge Hankamer

**Title**

The Role of Morphological and Phonological Factors in Bulgarian Allomorph Selection

**Permalink**

https://escholarship.org/uc/item/7473d2rt

**Authors**

Harizanov, Boris
Gribanova, Vera

**Publication Date**

2011-09-08

# THE ROLE OF MORPHOLOGICAL AND PHONOLOGICAL FACTORS IN BULGARIAN ALLOMORPH SELECTION*

BORIS HARIZANOV
*University of California, Santa Cruz*

VERA GRIBANOVA
*Stanford University*

## 1 Introduction

Discussions about the morphosyntactic and morphophonological status of the Bulgarian definite article (DEF) have played a significant role in the development of theories of syntax-morphology interactions (Mayer 1987, 1988; Sadock 1991; Popova 2000; Franks 2001; Embick and Noyer 2001; Dost and Gribanova 2006, i.a.). This may be in part due to the 'mixed' properties of DEF, whose morphophonology is akin to that of word-internal elements, but whose placement within the nominal phrase and with respect to other nominal clitics is governed by what can appear to be syntactic factors.

The morphophonological facts—for example, the fact that the attachment of DEF bleeds word-final devoicing, while the attachment of clausal clitics does not (1)—are typically invoked to argue that DEF should be treated as a form of inflection (i.e., as the affixal reflex of syntactic features, realized on a word within the relevant domain) (Franks 2001; Dost and Gribanova 2006).

(1)   DEF bleeds word-final devoicing:

    a.   /bratovčed/ → [bratofčet] 'cousin'

    b.   /bratovčed + a/ → [bratofčeda]  'the cousin'

    c.   /bratovčed + $i_{pl}$/ → [bratofčedi] 'cousins'

    d.   /bratovčed+ $e_{aux}$/ → [bratofčet#e] 'it's a cousin'

On the other hand, such data require special attention in the context of syntactic and post-syntactic accounts (Dimitrova-Vulchanova and Giusti 1998; Embick and Noyer 2001). This is because DEF has the 'word-internal' morphophonological behavior of suffixes (compare (1b) and (1c)), but the same distribution as nominal clitics, whose phonology is 'word-external' (Tomić 1996; Franks 2001, i.a.).

Empirical conclusions about DEF's status—and, relatedly, its bearing on differing theories of morphosyntax—are based, with varying degrees of detail, on claims of the sort illustrated above for word-final devoicing. In this squib, we explore a somewhat less clear-cut empirical generalization, namely the complicated question of what conditions the allomorphic form of DEF.

The form of DEF can preliminarily be characterized as conditioned by what appear to be both phonological and morphosyntactic factors (Scatton 1984; Caink 2000; Franks 2001). It can be demonstrated that in the case of some forms, 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.

(2) a. gaz (masc) 'gas (scientific term)' → gaz**a** 'the gas'

    b. gaz (fem) 'gas (other)' → gaz**ta** 'the gas'

(3) a. med (masc) 'honey' → med**a** 'the honey'

    b. med (fem) 'copper' → med**ta** 'the copper'

(4) a. prax (masc) 'powder' → prax**a** 'the powder'

    b. prax (fem) 'dust' → prax**ta** 'the dust'

(5) a. prəst (masc) 'finger' → prəst**a** 'the finger'

    b. prəst (fem) 'soil' → prəst**ta** 'the soil'

On the other hand, the phonological shape of the host noun can also be shown to play a role. The forms below show that there are two possible noun-final vowels in the plural forms of certain nouns, which are in free variation. When DEF is attached to such forms, its vowel must match the vowel immediately preceding it (inside the host).[1]

(6) a. kolen**a** 'knees' → kolen**ata** 'the knees'

    b. kolen**e** 'knees' → kolen**ete** 'the knees'

(7) a. ramen**a** 'shoulders' → ramen**ata** 'the shoulders'

    b. ramen**e** 'shoulders' → ramen**ete** 'the shoulders'

Any generalization about the allomorphic selection of DEF, then, will need to refer to both phonological (final segment) and morphosyntactic factors (gender and number). While this has been generally acknowledged, we know of no serious attempt to arrive at a comprehensive picture of DEF's various forms or of the conditions under which they arise. Here, we formulate those conditions (§2), and further attempt to draw from those generalizations some conclusions about their consequences for theories of morphosyntax-morphophonology interactions (section 3). In particular, the emerging empirical picture suggests that the mechanism governing allomorph selection must have access to both morphosyntactic and phonological features.

## 2 Toward a Generalization

Relying on the data presented below, we identify five distinct allomorphs of DEF in Bulgarian, whose attachment locales can be summarized as follows:

(8) a. *-a*: consonant-final masculine nouns

    b. *-tá*: consonant-final feminine nouns (the accent mark indicates stress shift from the stem)

    c. *-to*: neuter nouns

    d. *-te*: plural nouns

    e. *-tV*: nouns that end in V, where V is *-a* or *-o*[2]

We hypothesize that allomorph selection follows the following informal algorithmic statement:

(9) If the stem ends in V where V={*-a*, *-o*}, DEF → *-tV*;
    otherwise,

---

[1]In later sections of this squib we propose that phonologically conditioned DEF can be found on singular nouns as well.

[2]This informal insertion rule is stated in terms of vowel copying. The careful reader will have noticed, from examples (6) and (7), that there may be an alternative analysis of these data that appeals to vowel harmony. This would be inconsistent with the general phonology of Bulgarian, however, which makes no use of vowel harmony. It would also be inconsistent with the neuter forms in (15b,c) and the plural forms in (16a,c).

    a.  if singular masculine, DEF → *-a*

    b.  if singular feminine, DEF → *-tá*

    c.  if singular neuter, DEF → *-to*

    d.  if plural, DEF → *-te*

In other words, the choice of allomorphs (8a)-(8d) is based on the number and gender features of the head noun while the realization of allomorph(s) (8e) is phonologically conditioned.[3]

## 2.1 Masculine Forms

The majority of masculine nouns are consonant-final and take the *-a* allomorph of DEF.

(10)    Consonant-final masculines:
      məʒ**a** 'the man'
      brat**a** 'the brother'
      stol**a** 'the chair'
      prozorec**a** 'the window'
      luk**a** 'the onion'
      čajnik**a** 'the teapot'

Historically, a distinction between the "short" (-a) and "long" (-ət) form of DEF has been drawn, but since this distinction is largely artificial and the result of prescriptive pressures (Mayer 1988), we are not concerned with it here. The small set of vowel-final masculine nouns includes some kinship terms, certain occupational nouns, and month names.

(11)    Vowel-final masculines:

    a.  *-a*
        baʃta**ta** 'the father'
        vojvoda**ta** 'the voivode'

    b.  *-j͡a*
        sədija**ta** 'the judge'
        bojadʒija**ta** 'the painter'

    c.  *-o*
        tatko**to** 'the father'
        djado**to** 'the grandfather'

    d.  *-e*
        (seems to be a gap)

    e.  *-i* (most months and nothing else)[4]
        *januari-{a, ta, to, te} 'the January'
        *avgust-{a, ta, to, te} 'the August'

---

[3]A natural way to test this generalization would be to probe native speakers' intuitions about allomorph selection for DEF in a production study with nonce words.

[4]The impossibility of attaching DEF to month names is likely to be a lexical gap, analogous to the situation for certain kinship terms (Embick and Noyer 2001). This is supported by cases in which DEF surfaces in the presence of an adjective modifying the month name:

(i)    studenij-**a** januari
        cold-DEF  January.MASC

        'the cold January'

(ii)    toplij-**a**   juni
         warm-DEF June.MASC

         'the warm June'

To summarize, masculine nouns that end in a vowel take the allomorph of DEF that contains that final vowel (*-ta*, *-to*), while those that end in a consonant take *-a* (see (8a) and (8e)).

## *2.2 Feminine Forms*

Most feminine nouns end in *-a* or *-ĵa* and take the *-ta* allomorph of DEF, as the following examples illustrate:

(12)   Vowel-final feminines:

   a.   *-a*
        ʒena**ta** 'the woman'
        voda**ta** 'the water'
        rodina**ta** 'the motherland'
   b.   *-ĵa*
        zemja**ta** 'the earth'
        staja**ta** 'the room'
        ideja**ta** 'the idea'

There is a class of feminine nouns which end in *-e/ost* and take *-ta* as DEF. However, in the resulting definite forms, the stress always moves to the syllable of DEF:

(13)   Consonant-final feminines:

   a.   *-ost*
        mladost**tá** 'the youth'
        xubost**tá** 'the prettiness'
   b.   *-est*
        doblest**tá** 'the valour'
        prelest**tá** 'the delight'

Finally, there is an additional class of feminine nouns which end in a consonant and take this same stress-attracting *-ta* allomorph of DEF:

(14)   Other consonant-final feminines:

   a.   cev**tá** 'the barrel'
   b.   skrəb**tá** 'the grief'
   c.   gibel**tá** 'the destruction'
   d.   večer**tá** 'the evening'
   e.   prolet**tá** 'the spring'
   f.   smərt**tá** 'the death'
   g.   ljubov**tá** 'the love'
   h.   pesen**tá** 'the song'

In sum, feminine nouns that that end in a vowel take the allomorph of DEF that contains that final vowel (*-a*), while those that end in a consonant, take the stress-attracting *-tá* (see (8b) and (8e)). It may appear from these data that *-ta* is gender-conditioned, with stress shift being a property of the stem. We will argue against this view in §2.6.

*2.3 Neuter Forms*

Neuter nouns in Bulgarian all end in a vowel and take the *-to* allomorph of DEF (see (8c)). Words that end in *-i* and *-u* or *-jû* are loanwords, while those ending in *-o* and *-e* are native words.

(15)  Neuters: always vowel-final

    a.  *-o*
       selo**to** 'the village'
       dərvo**to** 'the tree'

    b.  *-e*
       dete**to** 'the child'
       more**to** 'the sea'

    c.  *-i, -u, -jû (loanwords)*
       taksi**to** 'the taxi'
       biʒu**to** 'the jewel'
       menju**to** 'the menu'

*2.4 Plural Forms*

Finally, in the plural, where no gender distinctions are maintained, two common noun endings are *-i* and *-ove*. The plural nouns that bear these take the *-te* allomorph of DEF.

(16)  *-i, -e*:

    a.  narod-i-**te**
       people-PL-DEF
       'the peoples'

    b.  məʒ-e-**te**
       man-PL-DEF
       'the men'

    c.  ʒen-i-**te**
       woman-PL-DEF
       'the women'

Additionally, there are a few pluralizing suffixes which all end in *-a* or *-jâ* and take the *-ta* allomorph:

(17)  *-a, -jâ*:

    a.  brat-ja-**ta**
       brother-PL-DEF
       'the brothers'

    b.  krai-ʃta-**ta**
       end-PL-DEF
       'the ends'

    c.  more-ta-**ta**
       sea-PL-DEF
       'the seas'

In sum, plural nouns that that end in *-a* take the *-ta* allomorph of DEF, while the rest take *-te* (see (8d) and (8e)).

*2.5 A Note on Noun Modifiers*

So far we have only looked at DEF attaching to nouns. However, the generalization expressed in (9) also captures the behavior of DEF allomorph selection in the presence of adjectival (and other) nominal modifiers. In nominal phrases that contain such modifiers, DEF attaches to the left-most modifier head. Because of DP-internal concord, all modifiers carry the number and gender feature of the head noun and agreement is always spelled out on modifiers as one of four things:

(18)   Number/gender adjectival suffixes:

    a.   *-∅*: masculine singular

    b.   *-a*: feminine singular

    c.   *-o*: neuter singular

    d.   *-i*: plural

    Now, as the allomorph selection algorithm predicts, masculine adjectives (and other agreeing modifiers) take *-a* as the definite article, feminine and neuter ones take *-tV*, where V matches the preceding vowel, and plural ones take *-te*:

(19)   a.   novij-a film
          new.MASC-DEF movie.MASC.SG

      b.   nova-ta kniga
          new.FEM-DEF book.FEM.SG

      c.   novo-to menju
          new.NEUT-DEF menu.NEUT.SG

      d.   novi-te knigi
          new.PL-DEF book.PL

That vowel matching, rather than gender matching, is the relevant generalization is supported by the fact that the stress-attracting *-tá* is never realized on nominal modifiers.

*2.6 Summary*

The data presented above point to the existence of the five allomorphs of DEF in Bulgarian as summarized in (8) and their distribution is captured by the generalization in (9). Note that this algorithm makes the correct predictions with respect to the "multiple plurals" cases (*kolena* vs. *kolene*) and the homonym pairs discussed in the previous section.

    A natural alternative to this analysis of the data might limit the distribution of phonologically conditioned DEF allomorphs to the masculine form, while treating the allomorphs in all other genders and numbers as morphosyntactically conditioned. This would mean that *-ta* and *-to* endings are phonologically conditioned allomorphs of DEF in the masculine, dependent on the final vowel, as is the *-a* ending (in consonant-final stems). Feminine and neuter nouns would take *-ta* and *-to* respectively on the basis of gender features alone.

    The difficulty with this view arises when we look at the plural forms. Plurals have a morphosyntactically conditioned allomorph, *-te*, whose distribution is in accordance with the elsewhere condition. However, *-ta* appears on plurals only if the plural form ends in *-a*. This suggests that unstressed *-ta*, where it appears, is a reflex of purely phonological conditions. The case for *-to* as solely phonologically conditioned is somewhat weaker since only masculine forms that end in in *-o* also take *-to*. This last analytical decision does not play a role in the discussion in §3, so we leave the question open.
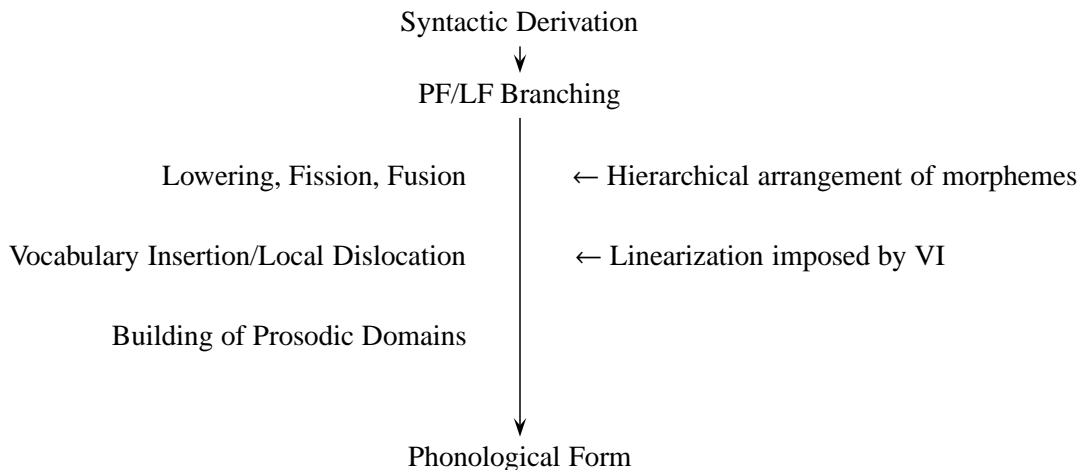
# 3 Allomorphy and Morphology-Phonology Interactions

A fundamental question for theories of allomorphic selection is what kind of information they should have access to. The empirical generalization arrived at in §2 indicates that DEF's form will be determined, in a specific phonological environment, by the phonological properties of its nominal host; otherwise, gender and number features will be the determining factor. A priori, there are two approaches to such generalizations. A first approach posits simultaneous access to both local morphological (gender, number) information and the phonological form of DEF's host. A second possibility involves a more restricted model, in which the process responsible for allomorph selection has access only to morphological features, and feeds phonological processes (which can no longer refer to morphological features).

Here, we implement these two approaches in the framework of Distributed Morphology (Halle and Marantz 1993, 1994; Embick and Noyer 2001; Embick 2010, i.a.), manipulating only the part of the theory that dictates the degree of access to different kinds of information at the point of allomorph selection. That the discussion is couched in terms of a piece-wise theory of morphosyntax-morphophonology interactions should be viewed as a matter of convenience, since the broad question will likely arise under any theory that aims to account for such data.

## 3.1 A Classical DM Approach

In the framework of DM, morphology interprets the output of narrow syntax where the atoms of syntactic computation are abstract feature bundles (*morphemes*). A number of strictly and universally ordered operations apply to a representation from the point of PF/LF branching to the point of Spell-Out (see figure 1). These operations account for the lack of isomorphism between syntax and morphology and exhibit sensitivity to hierarchical structure. At Vocabulary Insertion, syntactico-semantic features are replaced by phonological content and linear precedence relations are imposed. A Vocabulary item (VI) is the relation between a phonological string and information about where it can be inserted. Further displacement operations manipulate the linearized string and can only refer to string-adjacent elements (Local Dislocation) and phonological/prosodic information (Prosodic Inversion).

Syntactic Derivation
↓
PF/LF Branching

Lowering, Fission, Fusion     ← Hierarchical arrangement of morphemes

Vocabulary Insertion/Local Dislocation     ← Linearization imposed by VI

Building of Prosodic Domains
↓
Phonological Form

**Figure 1:** Operations in PF (based on Embick and Noyer 2001: *Fig.* 1)

(20)    VI schema

SIGNAL ↔ MORPHEME in CONTEXT

The terminal morpho-syntactic node that gets spelled out as DEF in Bulgarian is a feature bundle of the form exemplified in (21), which acquires its number and gender feature values via agreement with the head noun.

(21)    Morpho-syntactic representation of DEF

[def, number, gender]

Consider the four possible VIs that can be inserted as DEF, depending on the number and gender features of the morpheme:

(22)    a.    [def, sg, masc] ↔ /-a/
        b.    [def, sg, fem] ↔ /-tá/
        c.    [def, sg, neut] ↔ /-to/
        d.    [def, pl] ↔ /-te/

Insertion of these exponents captures the pattern only partially, since we know that DEF will be spelled out as /-ta, -to/ if its host ends in /a, o/ regardless of the number and gender of the head noun. In the classical conception of Vocabulary Insertion (see (20)), VIs cannot make reference to phonological context. In cases where the information in the VIs is insufficient to ensure generation of the correct output, the only option that remains for a standard DM model is to make use of a set of readjustment rules to obtain the correct phonological output (cf. the distinction between morphophonemic alternations and suppletion/conditioned allomorphy).[5] The readjustment rules that drive DEF allomorphy must be the following:

(23)    DEF → ta/STEM/a/__
(24)    DEF → to/STEM/o/__

That the resort to readjustment rules is necessary in this context is worrying for two reasons. First, readjustment rules, and the need for them, raises worries because these rules are by their nature largely unconstrained and non-local (Harley and Noyer 1999; Siddiqi 2006). A second reason to be concerned in this particular case is that the triggering environment for the application of a readjustment rule is supposed to be a morphosyntactic, not a phonological, one (Halle and Marantz 1993; Embick 2010). It seems that the classical DM view, in which Vocabulary Insertion has no access to phonological information in the insertion environment, is by its nature too restrictive.

## 3.2 An Alternative DM Approach

To implement in DM the approach in which allomorph selection has simultaneous access to both morphological information and the phonological form of DEF's host, one change in the set of assumptions needs to be made. The contextual information specified in VI (i.e., the environment of insertion) must be able to refer to phonological information (see the discussion of the Seri passive in Embick 2010:3.4.1).

(25)    VI schema

SIGNAL ↔ MORPHEME in CONTEXT, where CONTEXT can be morphosyntactic or phonological context

---

[5]For example, the *break/broke* alternation is considered to result from the application of such a readjustment rule.

This type of allomorphy is subject to strict locality conditions, under which the contextual condition of the VI refers to the phonology of a linearly adjacent element. Though we do not propose an analysis here, the basic requirement for a morphosyntactic analysis of DEF is that the phonological form of the nominal host be realized at the point that DEF is inserted. This in turn will depend on how locality interacts with postsyntactic rules like Lowering (Embick and Noyer 2001) and Clitic Metathesis (Arregi and Nevins 2008 and Harizanov, this volume).

With this amendment in place, we can view the Bulgarian DEF allomorphy as competition for selection among both gender-based and phonology-based VIs:

(26)  a. [def, sg, masc] ↔ /-a/

b. [def, sg, fem] ↔ /-tá/

c. [def, sg, neut] ↔ /-to/

d. [def, pl] ↔ /-te/[6]

e. /-a/[def] ↔ /-ta/

f. /-o/[def] ↔ /-to/

VIs that compete for insertion are subject to the Subset Principle (Halle 1997), which posits that the most specific VI (i.e., the one that matches the most feature of a morpheme) will be inserted:

(27)  Subset Principle

The phonological exponent of a Vocabulary item is inserted into a morpheme in the terminal string if the item matches all or a subset of the features specified in the terminal morpheme. Insertion does not take place if the Vocabulary item contains features not present in the morpheme. Where several Vocabulary items meet the conditions for insertion, the item matching the greatest number of features specified in the terminal morpheme must be chosen.

Given the featural content of DEF in (21), unless additional concessions are made, the Subset Principle will predict that the phonologically conditioned allomorphs (26e,f) will always lose the competition for insertion (because they are less specific), contrary to fact. It is not enough, then, to assume a theory in which there is unqualified simultaneous access to phonological context and morphosyntactic features.

If we are to maintain that a form of the Paninian principle is still operative here, an extension of the Subset Principle seems necessary. Vocabulary Insertion must identify the presence in a VI of a contextual condition referring to phonology as the most "highly specific" and "complex" environment for insertion (regardless of number of features matched, for example).[7] The Bulgarian data, however, point to the necessity of an additional assumption: phonological contexts must be a priori more "specific" with respect to the subset principle than morphological contexts.

## 4 Conclusion

Two sets of conclusions emerge from the present discussion. The first is an empirical one: the form of DEF must be stated as the result of both morphosyntactic and phonological properties of its host. The second is a theoretical conclusion: the mechanism governing allomorph selection must have access to both

---

[6]In order to arrive at the correct form for DEF for plural hosts, this algorithm must have access to the plural phonological form of a noun. This follows from a system that assumes, following the standard theory, that Vocabulary Insertion proceeds from the bottom up, in combination with the not unreasonable assumption that number information is realized as an abstract feature hosted on the root.

[7]Note that this is not equivalent to Hankamer and Mikkelsen's (2005) Extended Subset Principle which, in the presence of feature ties, favors the VI with the most specific morphosyntactic context for insertion. We leave for future research the question of the optimal definition of specificity, as well as the question of how these distinct notions should interact with one another, if at all.

morphosyntactic and phonological features. Further, if allomorphs compete for selection according to the Paninian principle, the phonological context must count as the more specific one.

# References

Arregi, Karlos, and Andrew Nevins. 2008. A principled order to postsyntactic operations. Ms. (lingBuzz/000646).

Caink, Andrew. 2000. In favour of a 'clitic cluster' in the Bulgarian and Macedonian DP. In *Papers from the 3$^{rd}$ Conference on Formal Approaches to South Slavic and Balkan Languages*, ed. Mila Dimitrova-Vulchanova, Iliyana Krapova, and Lars Hellan, volume 34 of *University of Trondheim Working Papers in Linguistics*, 170–182.

Dimitrova-Vulchanova, Mila, and Giuliana Giusti. 1998. Fragments of Balkan nominal structure. In *Possessors, predicates and movement in the determiner phrase*, ed. Artemis Alexiadou and Chris Wilder, 333–360. John Benjamins.

Dost, Ascander, and Vera Gribanova. 2006. Definiteness marking in the Bulgarian. In *Proceedings of the 25th West Coast Conference on Formal Linguistics*, ed. Donal Baumer, David Montero, and Michael Scanlon, 132–140. Somerville, MA: Cascadilla Proceedings Project.

Embick, David. 2010. *Localism versus Globalism in Morphology and Phonology*. Cambridge: MIT Press.

Embick, David, and Rolf Noyer. 2001. Movement operations after syntax. *Linguistic Inquiry* 32:555–595.

Franks, Steven. 2001. The internal structure of Slavic NPs, with special reference to Bulgarian. In *Generative Linguistics in Poland: Syntax and Morphosyntax*, ed. Adam Przepiórkowski and Piotr Bański, 53–69.

Halle, Morris. 1997. Distributed Morphology: Impoverishment and fission. In *Papers at the Interface*, ed. Yoonjung Kang Benjamin Bruening and Martha McGinnis, MIT Working Papers in Linguistics 30. Cambridge, Mass.: MIT Department of Linguistics and Philosophy, MITWPL.

Halle, Morris, and Alec Marantz. 1993. Distributed morphology and the pieces of inflection. In *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger*, ed. Kenneth Hale and Samuel Jay Keyser, 111–176. Cambridge, MA: MIT Press.

Halle, Morris, and Alec Marantz. 1994. Some key features of distributed morphology. *MIT Working Papers in Linguistics* 21:275–288.

Hankamer, Jorge, and Line Mikkelsen. 2005. When movement must be blocked: A reply to Embick and Noyer. *Linguistic Inquiry* 36:85–125.

Harley, Heidi, and Rolf Noyer. 1999. Distributed Morphology (State-of-the-article). *Glot International* 4:3–9.

Mayer, Gerald. 1987. The morphological categorization of the Bulgarian definite article. *International Journal of Slavic Linguistics and Poetics* XXXV-XXXVI:143–150.

Mayer, Gerald L. 1988. *The Definite Article in Contemporary Standard Bulgarian*. Wiesbaden: Otto Harrassowitz.

Popova, Guergana. 2000. Towards an HPSG account of the Bulgarian definite article. In *Proceedings of the ESSLLI-2000 Student Session*, ed. Catherine Pilière, 220–231. Birmingham, UK.

Sadock, Jerrold M. 1991. *Autolexical Syntax: A Theory of Parallel Grammatical Representations*. Chicago: The University of Chicago Press.

Scatton, Ernest. 1984. *A Reference Grammar of Modern Bulgarian*. Columbus, OH: Slavica.

Siddiqi, Daniel. 2006. Minimize Exponence: Economy Effects on the Morphosyntactic Component of the Grammar. Doctoral Dissertation, University of Arizona.

Tomić, Olga Mišeska. 1996. The Balkan Slavic clausal clitics. *Natural Language & Linguistic Theory* 14:811–872.