THEMATIC INTONATION PATTERNS IN BULGARIAN CLITIC REPLICATION

Tania Avgustinova
Department of Computational Linguistics
University of Saarland

Bistra Andreeva
Institute of Phonetics
University of Saarland

ABSTRACT

The phenomenon of clitic replication in Bulgarian involves two linguistic entities: a full-fledged nominal constituent in a particular syntactic function and a co-referent weak pronominal form corresponding to the particular syntactic function. In this paper we present experimental results which support our hypothesis that the nominal material that is replicated by a pronominal clitic in Bulgarian utterances consistently exhibits thematic intonational properties.

1. WORD ORDER TYPES AND INFORMATION STRUCTURE

Let us first introduce some background ideas and terminology concerning word order and information structure. In languages where the word order is discourse-conditioned and reflects to a considerable extent the information structure of utterances, i.e. the so-called free word order languages, a general (and fairly universal) distinction can be made between objective and subjective alignment. Presented in terms of the traditional theme-rheme partitioning of the sentence, an objective alignment (OA) would presuppose that thematic parts of the utterance linearly precede the rhematic ones, i.e.

$$[\text{theme} +] < [\text{rheme} +]$$

A subjective alignment (SA), in turn, amounts to a theme-preceding-rheme linear order, i.e.

$$[\text{rheme} +] < [\text{theme} +].$$

These two notions serve as basis for determining the *communicatively unmarked* - and *communicatively marked* - word order, respectively, depending on the utterance type. In particular, subjective alignment is considered unmarked for Bulgarian interrogative, imperative and exclamative utterances, with objective alignment pattern being the marked word-order variant. As for Bulgarian declarative utterances, communicatively unmarked word order conforms to the objective alignment, while the subjective one is communicatively marked. The importance of intonation for *marked* word orders in Bulgarian has already been observed by Ivanchev (cf. [8]).

Since we need to distinguish in our study between two types of thematic information - a more prominent discourse-anchoring part and a rather backgrounding "explanatory" part, we shall adopt the terminology used in the Information Packaging approach by Engdahl and Vallduví ([6], [9]). The basic focus-ground articulation of the utterance is further refined by dividing the ground into link (what the focus is about) and tail (how the focus fits in the context). Interpreted in these terms, an objective alignment would amount to a link-preceding-focus order, i.e. GROUND (link) < FOCUS, and the focus-preceding-tail alignment, i.e. FOCUS < GROUND (tail), is to be considered as subjective.

Linguistic resources relevant for the information structure (IS) of Bulgarian utterances involve:
- *word order*, remarkably flexible and discourse conditioned, as in all Slavic languages;
- morphological category of *definiteness*, unusual in the Slavic language family;
- *cliticisation*, as in any language having a system of weak and strong forms of personal pronouns: entities that are fully recoverable from the context can be structured as insignificant for the current communication purpose by using a clitic, an informationally inert element occurring in the utterance for reasons of (morpho-)syntactic well-formedness;
- *clitic replication* of nominal material, specific to Bulgarian;
- *intonation*, fairly malleable, as in languages like English and unlike, e.g., Czech.

From such a multidimensional perspective, Avgustinova models in [4] the IS of Bulgarian utterances as an interplay of three factors: the lexeme-specific obliqueness hierarchy of grammatical relations, the actually observable constituent order, and the contingent clitic replication.

Our goal in this article is to consider the intonational manifestation of the IS within task-oriented dialogues. We will be especially concerned with the intonation contours observable with nominal material which has been replicated by a clitic pronoun.

2. CLITIC REPLICATION

The grammatical process of manifesting as a clitic personal pronoun the INDEX (i.e. the person, number and gender) and the CASE (i.e. the syntactic function) of nominal material which is overtly realised in the same syntactic domain is referred to as clitic replication (CR). The phenomenon of CR involves two linguistic objects:
- a full-fledged nominal constituent - "base NP" - in a particular syntactic function (*direct object, indirect object, possessor*), and
- a co-referent weak pronominal form - "replica" - corresponding to this syntactic function (*accusative clitic, dative clitic, possessive clitic*).

The "base NP" (i.e. the "original") and its pronominal clitic

"replica" are related to each other through their grammatical categories (in particular, through agreement-relevant features).

Following [4], we distinguish two CR-relevant syntactic domains in Bulgarian - the clause and the NP, as well as two types of CR - anaphoric and cataphoric. Within the clause, both the direct and the indirect object NP are replicable by the corresponding dative or accusative verbal clitic (as illustrated in (1) and (3)), and in the syntactic domain of the NP, the possessor phrase can be replicated by a possessive clitic (as illustrated in (2) and (4). The CR is interpreted as *anaphoric* if the replicated nominal material linearly precedes the clitic within the relevant syntactic domain, as in (1) for the clause or (2) for the NP domain, and as *cataphoric* otherwise - cf. (3) or (4), respectively. For convenience, the nominal material and the clitic replicating it are uniformly underlined in the examples.[1]

(1) *Na nego programata mu ja pokazaxme nie.*
to him the-program DAT-3sg.m ACC-3sg.f showed we
(~ 'It was us who showed him the program.*')*
(2) *Na Ivan sestra mu e studentka.*
of John sister POSS-3sg.m is student
('John's sister is a student.')
(3) *Nie mu ja pokazaxme na nego programata.*
we DAT-3sg.m ACC-3sg.f showed to him the-program
('We showed him the program.'))
(4) *Sestra mu Na Ivan e studentka.*
sister POSS-3sg.m of John is student
('John's sister is a student.')

Even though the accusative CR in the syntactic domain of the clause may serve as a basic means for identification of the direct object when the verb inflection does not unambiguously reveal the subject, the dominant function of Bulgarian CR is object thematicisation (cf. [4] for details). Assuming that the nominal material that is replicated by a clitic pronoun belongs typically to a thematic (i.e. ground) part of the sentence, we could predict the intonation contour with it associated with the part of the utterance that is co-referent with the clitic pronoun. In our experiment, we have concentrated on this hypothesis.

3. MATERIAL

A modified version of the Edinburgh Map Task [1] was carried out in the Sofia variety of Bulgarian. Ten speakers were involved in the experiment, and each speaker participated in two map tasks. We examined the intonation contours observable in map task data which contained anaphoric or cataphoric instances of CR. The analysis was carried out using waves(tm) in conjunction with the "transcriber" script which is a part of the English ToBI [5]. The speech data was orthographically transcribed and intonationally labeled.

4. ACCENT PATTERNS

The intonation contour associated with the replicated nominal material depends on how the latter is interpreted in the IS of the utterance. The observed regularities can be formulated with respect to the CR type: anaphoric CR implies *link* interpretation of the replicated nominal material, while cataphoric CR imposes *tail* interpretation on the replicated nominal material. The corresponding accent patterns can preliminarily be classified as
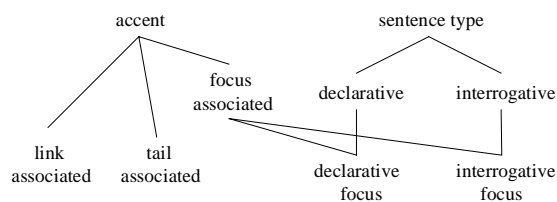
in Figure 1.

Figure 1

4.1. Link Associated Accent Patterns

Figure 2 illustrates the typically non-emphatic, non-contrastive rising F0 contour of a clause containing three link-associated accents and an anaphoric CR (here *go*) of the direct object (*starta*).

In the first and the second accent the post-accentual syllable (-*ta*, -*mash*) is on a higher pitch than the accented syllable (*star-*, *i-*). The pick in the third link-associated accent is towards the end of the accented syllable (-*chen*) because there are no more unaccented syllables following it.

The rise may begin at the left edge of the accented syllable (-*chen*), within the accented syllable (-*star*) or at its right edge (-*i*). The peak is reached within the post-accentual syllable, if there is any. According to the standard autosegmental approach these accents are represented phonologically as L*+H.
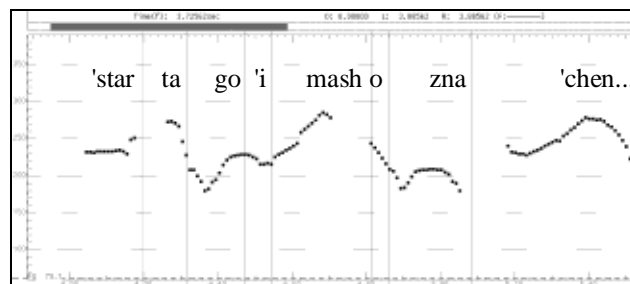
Figure 2 (female speaker ED)

[LINK **Starta go imash oznachen**] [FOCUS *malko vdjasno ot kashtichkata na muzeja.*]
the-start ACC-3sg.m you-have marked little right from the-house of the-museum
(~'Your start is marked a little bit to the right from the house of the museum.')

When two accents within the same intonational phrase are separated by only one unaccented syllable (cf. Figure 3 below) the H tone is reached at the end of the accented syllable (-*le*). This is an example of a configuration 'link–declarative focus–tail'. A more detailed discussion of the realization of focal accents in declarative and interrogative clauses can be found in [3] and [7].
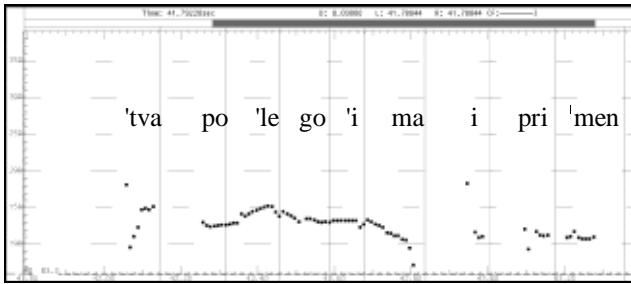
Figure 3 (male speaker ND)

[LINK *Tova pole*][FOCUS *go ima*] [TAIL *i pri men*].
　　this field   ACC-sg.n there-is   also by me
(~'I also have this field.')

Figure 4 illustrates a 'hat pattern' realization of the configuration 'link–focus'. That is, a rise (L*+H) on the direct-object nominal constituent in the link, followed by a plateau and a falling tone (H*) on the focused constituent – here a verb, and ending with a low boundary tone (L%).
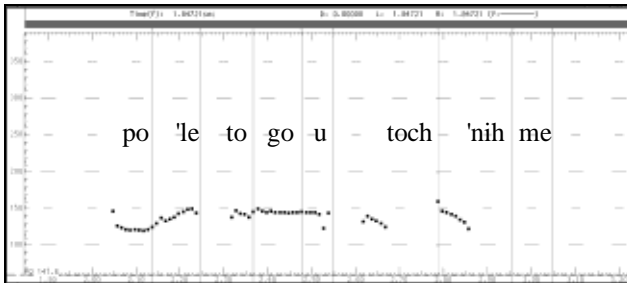

Figure 4 (male speaker ND)

[LINK *Poleto*] [FOCUS *go utochnihme*].
　　the-field   ACC-3sg.n we-localized
(~'We localized the field.')

Up to now, we have shown the intonational realizations of a replicated direct object. Figure 5 illustrates the link realization of an indirect object (*na tva*) anaphorically replicated by a dative clitic (here *mu*). There are three accents in this utterance. The first and the second are link-associated, while the third one is a focal accent. Similar to Figure 3, there is only one unstressed syllable (*mu*) between the two accented syllables (*tva*, *vi-*) which belong to the link part of the utterance, and the pitch movement is rising, with the peak being reached at the right edge of the accented syllable. The same can be observed with respect to the realization of the accentual pattern of the second link-associated accent, since there is just one unstressed syllable (*-kash*) between this accent and the focal accent.

　　This suggests that the intonation contour depends not only on the information structure but on the segmental material as well. Obviously, Bulgarian tends to realize bitonal accents on a single syllable whenever there is not enough material available.
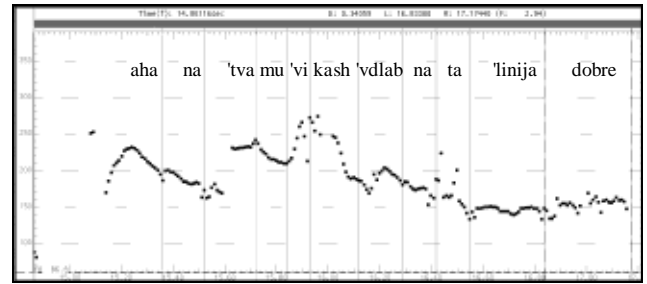

Figure 5 (female speaker MP)

Aha, [LINK *na tva mu vikash*] [FOCUS *vdlabnata*] [TAIL *linija.*] *Dobre.*
aha   to this DAT-3sg.n you-call   convex   line
(~'Ah, this is what you call a convex line. OK.')

In interrogative utterances the link is realized in the same way as in declaratives, i.e. by a rising F0 movement. What is the difference in the realization of the link and the interrogative focus, which is also realized with a rising movement?

　　Let us compare the link associated accent (*-na-*) and the focus-associated accent (*-lja-*) in Figure 6. Both accents are bitonal and rise from low to high. However, the H target of the focus is higher than the H target of the link. Moreover, the duration of the rising movement is longer in the case of a focus-associated accent, continuing on into the post-accented syllable. In Figure 3 and Figure 4 declination or 'hat pattern' can be observed. Figure 6, produced by the same male speaker and illustrating a sequence of link and focus in a query, demonstrates an inclination.
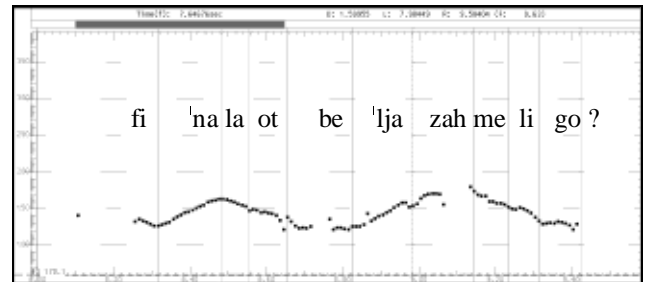

Figure 6 (male speaker ND)

[LINK *Finala*] [FOCUS *otbeljazaxme li go?*]
　　the-finish   we-marked   Q ACC-3sg.m
(~'Did we mark the finish?')

4.2. Tail Associated Accent Patterns

Note that we use the notion 'tail' in the sense presented in Section 1, i.e. designating a part of the utterance's information structure. It is prototypically realized as an intonational tail following the nucleus, which is also illustrated by the examples below. A systematic account of the intonational realization of the tail is given in [2].
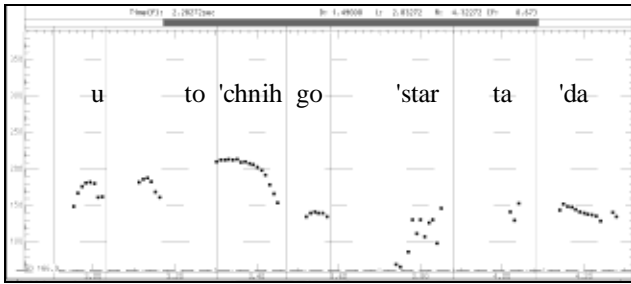
Figure 7 (female speaker PS)

[FOCUS *Utochnih go*] [TAIL *starta.*] *Da.*
    I-fixed ACC-3sg.m the-start   yes
(~'I fixed the start. Yes.')

In Figure 7 only a focus-associated accent (-*chnih*) is realized. The replicated material is in the tail, and is deaccented. The non-prominence of this material is phonetically expressed by a creaky voice. A tail realized in a query is shown in Figure 8. For a tail realization see also Figure 3 and Figure 5.

Figure 8 (female speaker EB)

[FOCUS *Ti li*] [TAIL *shte mi go   davash finala*]?
    you Q   will me ACC-3sg.m give the-finish
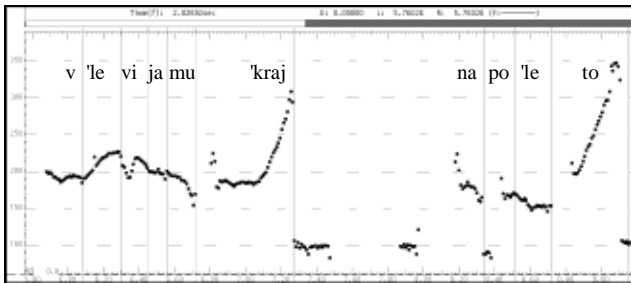(~'Is it you that will give me the finish?')

Figure 9 (female speaker ED)

I posle zavivash...tuka [FOCUS **v levija mu kraj**]...[TAIL **na poleto**]...
and then you-bend...here in the-left POSS-3.sg.n edge...of the-field
(~'And then you go around its left edge (of the field).')

In Figure 9, a replication of the possessor NP within the nominal syntactic domain is illustrated. The replicated material (*na poleto*) is realized as the tail. However, both the focal accent (*kraj*) and the tail, which has to be deaccented, are combined with a continuation rise. The conclusion to be drawn is that the continuation rise accent pattern represented with the boundary tone can override any other underlying accent pattern.

5. SUMMARY AND CONCLUSION

The results we obtained consistently support the hypothesis that replicated nominal material exhibits thematic intonational properties. The underlying (phonological) pitch accent pattern is L*+H. Differences in the particular phonetic realizations depend on how the theme is realized on the surface, i.e. as a link (non-final) or a tail (final).

If the replicated nominal material occupies a non-final position in the intonational phrase and has to be interpreted as a link in the communicative structure, the underlying pattern is realized phonetically as a gliding (slowly) rise in pitch from a low target within the accented syllable up to the next syllable (if there is enough syllabic material), otherwise only within the syllable itself.

If the replicated nominal material is in final position in the intonational phrase and is to be interpreted as a tail in the communicative structure, the underlying pattern is not realized phonetically, i.e. there is a phonological rule deleting all pitch accents after the nuclear tone.

NOTES

1.   In the glosses we shall use the following abbreviations:
ACC 'an accusative personal clitic pronoun', **DAT** 'a dative personal clitic pronoun', **POSS** 'a possessive personal clitic pronoun', **Q** 'the question particle', **1** 'the first person', **sg** 'the singular gender', **m** 'the masculine gender', **f** 'the feminine gender', **n** 'the neuter gender.

REFERENCES

[1] A.H. Anderson, M. Bader, E. G. Bard, G. Doherty, S.Garrod, S. Isard, J. Kowtko, J. Miller, C. Sotillo, H. Thompson and R.. Weinert 1991: The HCRC Map Task Corpus. In: *Language and Speech 34*. Vol. 4: 351-356

[2] B. Andreeva and T. Avgustinova: *On the Role of Intonation in the Information Structure of Bulgarian Utterances*. To appear as PHONUS 4. Research report, Institute of Phonetics, University of Saarland

[3] B. Andreeva and W. J. Barry 1997: Intonation von *checks* in der Sofia-Varietät des Bulgarischen. Presented at FDSL2 Potsdam (To appear in *Proceedings of FDSL2*)

[4] T. Avgustinova 1997: Word Order and Clitics in Bulgarian. PhD dissertation. Universität des Saarlandes, (Published in *Saarbrücken Dissertations in Computational Linguistics and Language Technology*, Vol.5, 1998)

[5] M. Beckman and G. Ayers 1994: *Guidelines for ToBI labelling, version 2.0.* Linguistic Department, Ohio State University

[6] E. Engdahl and E. Valldduví 1994: Information Packaging and Grammar Architecture: A Constraint-Based Approach. In: E. Engdahl (ed.) *Integrating Information Structure into Constraint-Based and Categorial Approaches* (DYANA-2 Report R1.3.B). Amsterdam: 41-79

[7] M. Grice, R. Benzmüller, M. Savino and B.Andreeva 1995: The Intonation of Queries and Checks across Languages: Data from Map Task Dialogues. In: *Proceedings of ICPhS'95*: 648-651

[8] S. Ivanchev 1957/1978: Nabljudenija vârxu upotrebata na chlena v bâlgarskija ezik. In: P. Pašov (ed.) *Pomagalo po bâlgarska morfologija. Imena*. Sofia: Nauka i izkustvo: 186-211

[9] E. Valldduví and E. Engdahl 1995: Information Packaging and Grammar Architecture: In: J. N. Beckman (ed.) *NELS 25*. Vol. 1. University of Pennsylvania: 519-533