

Word order in Topic-Focus structures in the Balkan languages

Iliyana Krapova

Introduction

Balkan languages can be said to belong to the so-called discourse-prominent languages, i.e. languages whose surface structure encodes through special syntactic means, rather than just prosodically, discourse(-semantic) functions such as Topic (discourse given or old information) and Focus (discourse new or emphatically represented information). In this contribution, I will show that at least in the three Balkan languages under study (Romanian, Bulgarian and Modern Greek, henceforth Greek), the two main types of discourse structures - topicalization and focalization – share a whole array of common syntactic properties and that their word order, at least in the preverbal field, is to a large extent shaped by information structure requirements.

We will use the terms ‘Topic’ and ‘Focus’, and we will speak of *topicalization* and *focalization*, respectively, since apart from providing a convenient methodology for empirical generalizations this precise theoretical way of capturing the role of information structure in syntax, has proved fruitful for syntactic description and is typologically well motivated. The terms themselves do not coincide with the traditional distinctions Theme and Rheme, although their essence captures the traditional Prague school intuition that each sentence can be divided into a discourse-familiar or discourse-given part (*theme*, *osnova*, *základ*) and a discourse-new part (*rheme*, *jádro* ‘nucleus’, cf. e.g. Cyxun 1962, Ivančev 1978).

Given the correlations between types of phrases in the preverbal field, as well as their relative order, the purpose of this contribution is to show that the so-called ‘Left Periphery’ (cf. Rizzi 1997) of the Balkan sentence is organized in a very similar way. Minimal variation between discourse structures is related to independent language internal differences, such as Case distinctions, the position of the clitic pronouns, use of special prepositions for object reduplication (such as *pe* in Romanian), etc.

While the existence of Left Peripheral structures is by no means an original Balkan phenomenon since it is present in diverse language groups from Romance to Semitic, the purpose of studying the Balkan Left Periphery is twofold. On the one hand, it can offer support for the presence of a universal Left Periphery, which has already been postulated typologically on the basis of a wide range of cross-linguistic studies; on the other hand, given that Topic and Focus structures are intimately related to purposes of communication and are most typical for colloquial speech, it comes as no surprise that the same mechanism underlying mutual comprehension could be held responsible for the quasi-identical ordering of phrases in sentence initial positions. One could also hypothesize that during the period when Balkanisms started to emerge, structures where discourse functions are overtly marked must have been favoured by speakers involved in any type of (bi- and multi-)lingual contacts (cf. Lindstedt 2000). While historical considerations will not play a role in the present contribution, this is nevertheless a direction worth being explored in the future.

Topic- and Focus-related notions can also be marked in sentence final positions. Both the sentence initial, and the sentence final positions (called ‘strong positions’ by Cyxun 1962, 268) are strongly endowed with discourse features, given discourse continuity. The sentence final position is typically associated with one type of Focus: New Information Focus (cf. Kiss 1998) or *rheme* in the strictest (classical) sense of the term. Following it, one can find also Topic elements (direct and indirect objects) which are typically marked by syntactic means

such as clitic doubling ('anticipatio', cf. e.g. Lopašov 1978)¹ or Clitic Right Dislocation (extraposition to the right). Cf. some examples from Bulgarian: *Ostavi ja onaja Marian - poznavam ja dobre; Toj izpālni tova, koeto mi obešta na men – da se bie mājki; Daj ja naj-setne taja legendarna posledna cigara; Az ne mu ja kazax istinata*. Such cases are in need of a better understanding not only in Bulgarian but in all of the Balkan languages given the pervasive use of such constructions (for a recent discussion on Greek and references, see Philiappaki-Warburton et al. 2004).

In this paper we will only concentrate on the Left Periphery, which following illuminating work by Rizzi (1997) has been applied to many languages in the last years and may therefore serve as a typologically well-motivated basis for any work on Balkan comparative syntax.

1. The Position of Topic and Focus in the Balkan Languages

The sentence-initial position of Topic and Focus is typical for all the Balkan languages under study.

(1) a. Ivan ne sām go viždala otdavna; b. **Samo Ivan** šte pokanja; Ivan, **nego** iskam da pokaniš.²

(2) a. Tin Eleni dhen tin idha; b. **To Jani** idhe i Maria; **Afton** thelo na kalesis.

(3) a. Pe Ion, I-am vāzut; b. **Mašinā** vrea Victor, nu casā.

Examples (1a, 2a, 3a) present topicalized direct objects; the examples in (1b), (2b), (3b) represent focalized direct objects.

Reduplication by a pronominal clitic (also referred to as 'reprisa') is the classical mark or Object Topicalization in all of the Balkan languages (Lopašov 1978, Cyxun 1981, Assenova 2002).³ From the point of view of current formal syntactic theorizing Topic structures are seen as involving **dislocation** of an (direct or indirect) object to the preverbal position. From the point of information structure, the word order corresponding to (1a), (2a), (3a), is referred to as *objective word order* ('prav slovoed', cf. Ivančev 1978), since sentence initial Topics are linked to the preceding discourse and thus serve as a starting point ('terme de départ', in Guéntcheva's 1994 terminology) for the actual predication. The Topic can also be viewed as the logical (notional) subject of the predication, i.e. what the predication is about. The rest of the sentence belongs to what is generally called 'Comment', i.e. the predication itself (cf. Vallduvi 1992). Since the clitic is obligatory, this type of dislocation has also been termed **Clitic Left Dislocation** (CLLD) (introduced by Cinque 1990 for similar constructions in Romance). This is the term we will be using here.

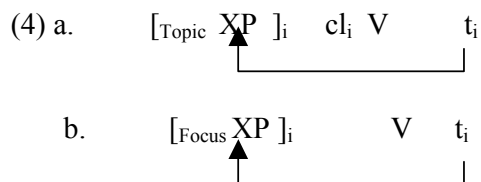
¹ We follow Lopašov (1978, 14) in differentiating two types of structures: those in which the object is preposed with respect to the verb (*reprisa*), and those in which the object is postposed (*anticipatio*). Although he considers the difference in quantitative terms, there are other, deeper, differences between these two structures. There are also historical considerations for such a distinction, at least in Bulgarian. As reported by Minčeva (1969), Topicalization qua preposing of the object is a much older phenomenon, as it can be found in a number of contexts in Old Church Slavonic. Typically, the anaphoric pronoun or a demonstrative pronoun used to double the preposed (heavy and intonationally independent) object. *Anticipatio*, on the other hand, is a later phenomenon – the earliest documents in which it is attested date from the 12th-13th c. According to Minčeva, the later expansion of *anticipatio*, while still attributable to the syntactic principles of colloquial speech, involves additional factors such as the position of the enclitic, the syntactic independence of the verbal group, etc.

² In all the examples to follow, focused phrases will be given in bold.

³ Historically, the primary function of the *reprisa* has been related to the grammaticalization of the SVO word order in the Balkan languages, following the loss of Case distinctions, whose most visible effects are observed in Bulgarian. Apart from ensuring a greater word order freedom and achieving discourse prominence, the topicalization of the object in a sentence initial position serves other syntactic purposes, such as the disambiguation of (potentially ambiguous) subject – object structures (cf. Lopašov 1978, 83, 99, 101-105, Assenova 2002, 108f), e.g. *Dimov go ubi Meri Lamour* (ex. from Popov 1962).

Focalized phrases, on the other hand, enter into another type of information structure articulation: the Focus - Presupposition articulation, well-known since Chomsky (1972). Sentence initial Focus (also referred to as Contrastive Focus or Identificational Focus, cf. Kiss 1998) is a specific type of Focus.⁴ Pragmatically, it expresses the speaker's intention to resolve a potential misunderstanding or doubt on the part of his interlocutor, or to correct some (part of a previous) statement. Therefore, Contrastive Focus is necessarily associated with some contextually determined set of alternatives for which the predicate holds potentially, by pointing out the unique member (or subset) of that set for which the predicate actually holds (Zubizarreta 1998, 6). Syntactically, this strategy makes use of the *subjective word order* ('obraten slovoed', cf. Ivančev 1978): the information is presented as the most relevant part of the utterance and is typically pronounced with (strong) emphasis, i.e. it carries emphatic stress ('logičesko udarenie', cf. Popov 1961, Cyxun 1962, 287). This type of Focus conveys new information only indirectly: by emphasizing the information the speaker typically brings forward a (potentially) novel quality or property of what is being talked about, i.e. of the discourse theme (Popov 1961).

Given the examples in (1b)-(3b), Focus can also be said to involve **dislocation**, but without an accompanying clitic pronoun. The dislocation of a Topic or a Focus to a preverbal position can be schematically represented as in (4), a & b respectively.



The abstract representations in (4) indicate that topicalization and focalization involve the same type of structure, differing only in the presence or absence of a clitic. In both cases the object starts out from an object position, as a verbal argument, and dislocates to the preverbal position, leaving a trace (t) in its original position. Only in (4a), the clitic mediates the syntactic relation between the preposed object and its trace, ensuring co-referentiality (Guéntcheva 1994, 119).

In the absence of doubling, i.e. when another type of phrase (prepositional phrase, adverbial phrase, etc.) preposes to a Topic or a Focus position, the difference between the two discourse structures is achieved only prosodically (low stress, flat intonation, intonational pause vs. emphatic stress). Naturally, in the absence of such clues, it is the context that resolves potential discourse ambiguities (cf. Joseph & Philippaki-Warbuton 1987, 99-102 for Greek, Rudin 1991 for Bg), cf. examples from Greek and Bulgarian:

(5) a. (Gr) Sto xorio tis pijeni poli sixna. Me se vazmo prepei na milate sto patera sas. Stu Jani na pame appose.⁵

b. (Bg) Predi njakolko dni beše xodila iz selo peperuda; Prez gorata, pravo kām mene, idexa kozi, Vāv vsjako xudožestveno proizvedenie trjabva da ima dviženie (AG 1994, 176), Na kino otivam (ne na sābranie).

⁴ This type of Focus should be strictly differentiated from New Information Focus, which, as mentioned above, corresponds best to the traditional notion of *rheme* and appears in a sentence final position, since it can be used as an answer to a question requesting new information, e.g. *Kakvo donese Ivan?* - *Ivan donese* [IF knjigite].

⁵ In Greek, dislocation for emphasis can be accompanied by an emphatic nonclitic proform, cf. *Stin Elada, eki na pame jia djakopes*; *Tin kiriaki, tote na pame* (Joseph & Philippaki-Warbuton 1987, 100)

1.1. Two types of topicalization structures

Having seen the basic structure types underlying topicalization and focalization, we proceed by noting that in the Balkan languages under study, two types of Topic structures can be distinguished. Thus, alongside (1a)-(3a), there exist cases like (6).

- (6) a. (Cît despre) Ion_i, l_i-am văzut **pe el_i** de anul trecut. (Rom)
b. (Kolkoto do) Ivan_i, včera go_i srešnax **nego_i**. (Bg)
c. (Oson afora tin) Maria_i, dhen tin_i anteho **afti_i** allo. (MG)

The constructions in (6) have been studied for each of the three Balkan languages (Dobrovie-Sorin 1990, 1994, Anagnostopoulou 1994, 1997, Rudin 1986, Džonova 2004, etc.). Our task here is to outline in a comparative way their cross-Balkan properties. The existence of the construction in (6) has been noted first for Romance (cf. Cinque 1977, 1990) and has been labeled **Hanging Topic Left Dislocation (HTLD)** – a term which was meant to distinguish it from CLLD. We will see below that the distinction between the two types of left dislocation is also valid for all the Balkan languages under study. In particular, Balkan HTLD shares all the properties characteristic of Romance HTLD, with one notable difference: while in Romance the resumptive element can be a tonic pronoun without any accompanying clitic, in the Balkan languages, the tonic pronoun must be doubled by a clitic, as the ungrammaticality of (7) compared with (6b) shows:

- (7) *(Kolkoto do) Ivan_i, srešnax **nego_i** včera.

In discussing cases like (6) in Bulgarian, Guentchéva (1994) argues that the extraposed term (the HT in our terminology) is co-referential only with (and reduplicated only by) the tonic pronoun. The real reduplication, however, takes place between the tonic pronoun and the clitic, since only this configuration is sentence-internal. Therefore, the author excludes the possibility of analyzing cases such as *A sãrceto# bjas go kãsa nego kletto* (p. 157) which are parallel to (6b) above, as involving “triple reduplication”.

1.2. Properties of Hanging Topics

Hanging Topics have clear pragmatic, prosodic and structural properties. First of all, from a pragmatic point of view, the relation of this type of Topic and the following Comment is rather loose, i.e. the HT creates only a general context for the Comment, which is why in the literature such constructions are also referred to in Guéntcheva (1994) and Assenova (2002) as *extraposition Topics, segmented phrases* (in the sense of Ch. Bally 1932/1965) or *thématisation forte* (‘strong Themes’). Additionally, from a prosodic point of view, there is a sharp intonational break between the left dislocated phrase and the rest of the sentence,⁶ especially if it is introduced by *as for* expressions (‘thématisateurs’ in Feuillet’s 1990 terminology), such as *što se otnasja do/kolkoto to, cît despre, oson ja/oson fora*, whose purpose of to clause off the HT from its Comment.

Despite these peculiarities of HTLD, which are not shared by CLLD, where the dislocated XP acts as a real double of the resumptive clitic and is necessarily interpreted in its base (argument) position, the two constructions are hard to distinguish when the dislocated Topic is a simple noun phrase, especially in the absence of an *as-for* expression or of a (sharp)

⁶ Following the standard practice, in the examples below the (heavy) intonational pause after the HT will be indicated with the symbol #.

intonational pause. Therefore, one needs to apply some other test from the range of diagnostics offered by Cinque's (1977) study on comparable Topic constructions in Romance.

The first and perhaps the most important diagnostic has to do with Case connectivity, i.e. the dislocated phrase and the resumptive element have to match in Case features (and not just in person/number/gender features). While the Topic construction referred to above as Clitic Left Dislocation meets the Case connectivity requirement, the Hanging Topic Left Dislocation does not. Only in this latter construction, the dislocated phrase can appear (and usually does appear) in the Nominative case (*Nominativus pendens*) rather than in the same Case as the resumptive element. Nominative Topics (resumed by a clitic in some other Case) are a clear instance of HTLD, as illustrated by the Greek example in (8) where the (obligatory) pause separates the initial Topic from the rest of the clause. Compare (8) with (6c) above where the Accusative Case of the initial Topic is required by the 'thematiseur' *oson afora* (Anagnostopoulou 1997, 154):

(8) I Maria# tin ematha kala tosa xronia, ksero pos na tis miliso.

Anagnostopoulou (1994, 1997) and Alexiadou (1997) show that in Greek (where the Nominative-Accusative distinction is preserved), Nominative Topics can only appear in root clauses. In embedded clauses, on the other hand, the dislocated object must have its regular Accusative case. This in itself already points to the fact that whenever we have the configuration $XP_i \dots cl_i$ in an embedded clause, we must be dealing with a CLLD structure, as in (9). Since there is a parallel restriction in Romance, we can view the obligatory root clause character of the HTLD as a second general diagnostics for the difference between this construction and the CLLD construction.

(9) Ipe oti *i Maria/ tin Maria tin emathe kala tosa xronia.

Other diagnostics prove crucial for Romanian and Bulgarian, given the absence of a Nominative vs. Accusative distinction in the nominal system of the former, and the absence of any Case distinctions in the nominal system of the latter. Carmen Dobrovie-Sorin (1990) suggests that in Romanian topicalized phrases introduced by *pe* can only enter the CLLD construction because this preposition (similarly to the preposition *a* in Spanish) is only licensed internally to the associated sentence with respect to a (definite and [+human]) direct object phrase. This proposal receives support from the incompatibility between dislocated objects introduced by *pe* and emphatic pronouns, which are typical for the HTLD.

(10) *Pe Maria nu vrea s-o mai väd **pe ea** cât trăiesc. (Dobrovie-Sorin 1990, 373)

The example thus illustrates the following generalization: any distinguishing property, which is compatible with just one of the two constructions and is found in a context compatible only with the other construction, yields ungrammaticality. We expect therefore that if a dislocated phrase resumed by a tonic pronoun, which is only compatible with HTLD construction, is found in an embedded context, ungrammaticality will arise, since embedded contexts are compatible only with the CLLD construction, but not with the HTLD construction. That this generalization is correct is shown by the ungrammaticality of (11a) from Bulgarian, which should be compared with the parallel case in (11b) featuring the CLLD construction:

(11)a. *Kazax, če Marija# az sām i kupil na neja cvetja.
b. Kazax, če na Marija az sām i kupil cvetja.

The sequence following the complementizer *če* in (11a) is grammatical if used in a root clause, cf. (12a). Also grammatical is the variant where the dislocated indirect object retains its preposition *na* ‘to’, cf. (12b). Note however, that the grammaticality of (12b) has to meet two additional criteria: no resumptive tonic pronoun is allowed to appear inside the clause and no intonational pause can follow the dislocated indirect object, cf. (12c) and (12d) which are ungrammatical because one or the other requirement is not met:

- (12) a. Marija# az sām i kupil na neja cvetja.⁷ HTLD
 b. Na Marija az sām i kupil cvetja. CLLD
 c. *Na Marija az sām i kupil na neja cvetja.
 d. *Na Marija# az sām i kupil cvetja.

What the Bulgarian examples reveal is another peculiarity of the HTLD construction: the dislocated phrase can only be a noun phrase (NP), not a prepositional phrase, nor a phrase of some other category. No such restriction exists for CLLD. This distinguishing property, offered as another diagnostic by Cinque (1977), is employed only in contemporary Bulgarian where indirect objects are prepositional phrases (PPs). Given that Romanian has no prepositional indirect objects and that in Greek, prepositional indirect objects cannot be clitic resumed (cf. *Sto Jani tha dhosi i Maria ta lefta avrio* - Brian & Philippaki-Warburton 1987, 99), then it must be the case that Topicalized Case marked indirect objects in these two languages may participate only in the CLLD construction,⁸ observing Case connectivity.

The diagnostics *presence/lack of Case connectivity effects* is applicable to Bulgarian only in the case of Topicalized pronouns which unlike nouns, show Case distinctions. As made evident recently by a corpus collected by Marina Džonova (cf. Džonova 2004), Bulgarian colloquial speech makes an abundant use of Nominative pronouns as left dislocated Topics. Two examples are given in (13). The absence of Case connectivity between the Topical pronoun *az* ‘I’ and the resumptive clitics *mi/me* ‘(to) me’ identifies the use of the HT strategy.

- (13) a. **Az#** na mene tova nikoga ne mi se e slučvalo.
 b. **Az#** mene me e jad, če si vključix Klip navremeto.

Both examples feature the tonic pronoun *mene* ‘me’ which, given its position after the intonational break, as well as Case connectivity effects, can only be analyzed as a CLLD object.⁹ Nominative Topics are also characteristic of (eastern) Bulgarian dialects and in fact, have been reported to exist from the earliest manuscripts reflecting in writing the specific properties of the colloquial language (13th c., cf. Minčeva 1969 for examples, references and discussion about the presumed archaic nature of such constructions). (14) below gives some dialectal examples, taken from Stoykov (1962/2002, 260) and Mladenov (1965, 213):

⁷ We do not mean that all criteria have to be met in order for a certain construction to qualify as a HT. For example, if a tonic pronoun is not realized in a certain structure, then Case connectivity becomes the distinguishing factor between a CLLD and a HT structure, cf. *Na Ivan otdavna ne sa mu plaštali vs. Ivan otdavna ne sa mu plaštali* (from Džonova’s corpus).

⁸ It could be the case that in Greek, dislocated indirect objects cannot function as HTs, cf. the ungrammaticality of (i) reported by Alexiadou (1997):

(i) *I Maria, o Janis tis ta edhose ta vivlia.

⁹ The tonic pronoun can also occur at the absolute end of the sentence. In this case, we are dealing with the mirror image of the CLLD – Clitic Right Dislocation. The latter is also typical for marking a Topical object (or as a kind of an afterthought) in Bulgarian, as well as in the other Balkan languages.

(14) a. As inn□' žinà beši mi kàzala (Belensko); b. Toj n'ama da gu ìma tà̀m (Slivensko); c. Ja ide mi se; d. Ja snošti ič mi se ne slizaše. (Ixtiman).

According to Mladenov (1965), Nominative pronouns as Topics are found even in dialects which do not allow clitic resumption, such as the Ixtiman dialect. However, we should be more careful in characterizing such constructions since there are further parametric differences between them which merit further research.

There is a further diagnostics in Bulgarian for differentiating CLLD Topics, namely the position of the clitic with respect to the dislocated object. As is well-known, differently from Romanian and Greek, Bulgarian clitics obey the Tobler-Mussafia law, i.e. they cannot occupy a first position after an intonational pause. Cf. the ungrammaticality of **Ivan# go vidjax nego včera*. According to Minčeva (1969, 19), cases in which the clitic leans on the last word of a previous phrase, as well as the inverse cases in which the clitic encliticizes without being related to its host, point to the fact that the position of the clitic is syntactic, rather than prosodic. In the Bulgarian data at hand, we observe that whenever there is no pause to separate the dislocated object from the rest of the sentence, the clitic is enclitic on this object, e.g. *Mene me čaka rabota*. This is the case of the CLLD construction. However, when the clitic follows after a pause, as it happens in the HTLD construction, either the verb inverts, and the clitic encliticize on it, as in (14c), or else, the clitic is hosted by an additional (CLLD) Topic or a Focus phrase, in preservation of the order Cl V, as in (13b), (14d). Given that according to Cyxun (1962) even V inversion around the clitic is informationally triggered, we can conclude that in the presence of an initial HTLD construction, the clitic can be hosted by whatever discourse material follows the HT.

However, we have to note that this is not always the case, since sometimes, in the absence of a pause, the clitic may encliticize on a Nominative Topic pronoun. This is frequent with experiencer constructions of the type *Az mi se iska*. Yet in other cases, we also find cases of a verb > clitic order following after a Nominative Topic, e.g. *Az# iskaše mi se da razbera nešto poveče za Tărnovo* (from Džonova's corpus). Such differences merit further research; here it is worth noting that Nominative Topics, at least with experiencer verbs, are not always HTs.

1.3. Linear orders

The example in (14) above from colloquial Bulgarian gives evidence that in case a HT co-occurs with a CLLD Topic, the former must precede the latter. As expected, the reverse sequence gives rise to ungrammaticality, whatever the intonational contour, cf. **Na mene az tova nikoga ne mi se e slučvalo*. This general property of HTs (namely, that they occupy an absolute sentence initial position) is supplemented by a uniqueness requirement: there can only be a single HT per sentence. CLLD Topics, on the other hand, are exempt from the uniqueness requirement. Consequently, more than one such Topic can appear per clause, and there is no particular order observed. The data collected by Alboiu (2000) for Romanian, by Anagnostopoulou (1994, 1997) for Greek, and by Arnaudova (2002) and Krapova (2002) for Bulgarian, confirm this generalization, although there seem to exist interpretational differences which need to be studied separately. So, for example, according to Alboiu (2000, 270), in Romanian, the highest Topic has maximum relevance for the discourse context, but otherwise all combinations are possible. In the examples below, Topics are given in brackets, so that their free ordering can be made more evident.

(15) a. [_TTa vivlia] [_Ttis Marias] tis ta edhose to Janis; [_TTis Marias] [_Tta vivlia] tis ta edhose o Janis.

b. [_TMioarei] [_Tinelul] la nuntă i l-a dat Anghel; [_TInelul] [_TMioarei] la nuntă i l-a dat Anghel.

c. [_TNa Marija] [_Tpismoto] i go dadox az; [_TPismoto] [_Tna Marija] i go dadox az.

1.4. *Movement of the CLLD object*

Recall that in section 1, we postulated that the Left Periphery of the sentence contains a Topic position which is targeted by clitic resumed material counting as a Topic. However, given the above discussion on the distinction between HTLD and CLLD, we should try to find out whether both types of Topicalization involve movement. Following the conclusions reached unanimously by all of the authors who have studied the distribution of HTs in the Balkan languages, we maintain that this particular type of dislocation is not derived by movement (cf. in particular Rudin 1986, Dobrovie-Sorin 1990, Anagnostopoulou 1997). Some arguments to this effect are presented below. As far as CLLD is concerned, Case Connectivity already indicates that movement has taken place: the matching clitic functions as an anaphoric element which connects the original (base) position of the dislocated argument to its surface position.

1.4.1. ‘Unboundedness’

A first piece of evidence that the CLLD construction is derived through movement comes from the fact that it is not limited to monoclausal domains (Anagnostopoulou 1997): the dislocated phrase can appear outside of the embedded clause to which it belongs. Hence the term ‘unboundedness’. (15) provides examples from Greek and Bulgarian showing that the embedded Topics have been dislocated into the domain of the matrix clause.

(16) a. Tin Eleni_i su ipa xthes oti t_i tin idha t_i.

b. Prestăpnika_i mislja, če t_i sa go xvanali t_i.

Such observations point to a movement operation - the Topic starts out from the complement clause and dislocates to a position in the Left Periphery of the embedded clause, after which it moves into the Left Periphery of the matrix clause. This is indicated by the identical indices on the traces left at the positions through which the Topic passes on its way to its surface position. We do have independent evidence that Topic has moved through a position to the right of the complementizer given the following variants of (16):

(17) a. Su ipa xthes oti [tin Eleni]_i tin idha t_i.

b. Mislja, če [prestăpnika]_i sa go xvanali t_i.

Topic Movement takes place also out of subjunctive complements and indirect questions, as illustrated by the following transformational pairs:

(18) a. Perimeno [ta lefta]_i na ta feri o Janis t_i → [Ta lefta]_i perimeno t_i na ta feri o Janis t_i.

b. Očakvam [parite]_i da mi gi donese Ivan t_i. → [Parite]_i očakvam t_i da mi gi donese Ivan t_i.

The examples above show that there is position to the left of the subjunctive complementizers (particles) *da/na*, as well as to the left of the *wh*-word in indirect questions, through which the Topic moves into before it continues to the matrix clause.

Hanging Topics are also unboundedly distant from their resumptive pronouns. However, differently from CLLD, they cannot appear in any intermediate position (given that they are

illegitimate in embedded clauses). Consequently, they are not moved from the embedded clause but are directly generated in the matrix clause.

1.4.2. *The position of anaphors*

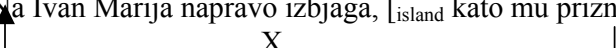
The second piece of evidence comes from the syntactic behavior of reflexive pronouns and expressions containing a reflexive pronoun. As is well known, such expressions function as anaphors which have to be bound by their antecedents. In all of the Balkan languages under study reflexives are impossible as HTs but are perfectly grammatical as CLLD Topics. Compare the following pairs:

- (19) a. *O eaftos tu_i # o Janis_i dhen ton frontizi t_i (Gr - Anagnostopoulou 1997,155)
 b. Ton eafto tu_i o Janis ton_i prostatevi t_i.
- (20) a. *Cît despre sine_i # Victor_i nu s_i-ar pune in pericol. (Rom - Alboiu 2000,272)
 b. Pe sine_i, Victor_i nu s_i-ar pune in pericol t_i.
- (21) a. *[Vsičkite si prijateli]_i # gledam da im pomogna t_i (s kakvoto moga). (Bg)
 b. [Na vsičkite si prijateli]_i gledam da im pomogna t_i (s kakvoto moga).

In all of the grammatical examples, the anaphor has to reconstruct to its base position (indicated by the trace) in order to be interpreted as bound by its antecedent which shares the same index. The ungrammatical examples, on the other hand, represent a reflexive contained within a HT. Since the anaphor is left unbound, we infer that no reconstruction has taken place. Therefore, such cases constitute evidence that the HT is generated directly in its surface position rather than moved there.

1.4.3. *Island sensitivity*

A third piece of evidence which distinguishes between presence vs. lack of movement has to do with islands. Islands are clauses (or phrases) that do not allow any phrase internal to them to move out. A typical example of islands is an adverbial clause (adjunct clause). The examples below are meant to show that HTs are not sensitive to any islands, because if they were, they would not be able to move out. CLLD, on the other hand, are sensitive to (strong) islands and therefore, movement out of the island is impossible (as indicated in (22b)):

- (22) a. (Kolkoto do) Ivan# Marija napravo izbjaga [_{island} kato go celuna].
 b. *Na Ivan Marija napravo izbjaga, [_{island} kato mu prizna vsičko].
- 

Similar data are reported for Romanian and Greek, examples (23)-(24):

- (23) a. (Cît despre) Ion # am plecat înainte să-l-examineaze Popescu; Cît despre Ion, n-am înfîlnit fata care l-a văzut ultima dată.
 b. *Pe Ion am plecat înainte să-l-examineaze Popescu; *Pe Ion n-am înfîlnit fata care l-a văzut anul trecut. (Dobrovie-Sorin 1994, 219)
- (24) *Tin efemerida apokimithike diavazontas. (Anagnostopoulou 1997, 172)

We summarize with the Table below all the properties of the two types of left dislocation constructions, and we add one more illustrating example.

	CLLD	HTLD
1. Case connectivity	yes ex. <i>Ivan/nego ne mogat da go prikrepjat kam nikogo.</i>	no ex. <i>Ti(#) ne mogat li da te prikrepjat kãm njakoj? Tja i bez tova ne moga da ja nakaram da jade.</i> (from Džonova's corpus of colloquial speech)
2. Tonic pronoun or a clitic pronoun	clitic <i>Ivan go čaka druga rabota.</i>	tonic + clitic <i>Ivan, nego go čaka druga rabota.</i>
3. Root or embedded clauses	Root and embedded clauses <i>Na Marija s ništo ne si ì pomognal.</i> <i>Ivan kaza, če na Marija s ništo ne si ì pomognal.</i>	Root clauses only <i>*Ivan kaza, če Marija# na neja s ništo ne si ì pomognal.</i>
4. Types of phrases	NP, PP, AdvP.... <i>Na Ivan otdavna na sa mu plaštali.</i> <i>Pismoto go napisax az.</i>	NP only <i>Ivan otdavna ne sa mu plaštali</i> (from Džonova's corpus)
5. Number of dislocated phrases	More than one <i>Tija knigi na vas koj vi gi e pratil ?</i> <i>Na vas tija knigi koj vi gi e pratil ?</i>	One <i>A ti # tebe xapalo li te e kuče?</i> (colloquial)
5. Strong islands Adjunct island	Sensitive <i>*Na Ivan Marija izbjaga, kato mu dade rozata.</i> <i>*Pe Ion am plecat înainte să-l examineaze Popescu.</i>	Not sensitive <i>Ivan# Marija izbjaga, kato mu dade rozata.</i> (Cît despre) <i>Ion, am plecat înainte să-l examineze Popescu.</i> <i>Oso ja to Jani, i Maria efige molis ton idhe.</i>
Complex NP (Relative Clause) island	Sensitive <i>*Na Ivan poznavas li onova momiče, koeto mu dava knigi?</i> <i>*Pe Ion n-am întilnit fata care l-a văzut anul trecut</i> <i>*To Jani dhen sinandhisa to koritsi pu ton ide.</i>	Not sensitive <i>Ivan# poznavas li onova momiče, koeto/deto mu dava knigi?I</i> (colloquial) (Cît despre) <i>Ion, n-am întilnit fata care l-a văzut ultima dată</i> <i>Afto to vivlio, ksero to singrafea pu to egrapse.</i>

1.5. Topic structures in embedded clauses

From the facts discussed so far the following empirical generalizations emerge:

- 1) Hanging Topics precede CLLD Topics in all of the languages under study;
- 2) Embedded CLLD Topics follow the declarative complementizers/subordinators *oti/če/că* (cf. example (25) from Romanian). Additionally, in Greek and Bulgarian CLLD Topics can sometimes (and for some speakers) appear in front of this complementizer, cf. (26).

(25) Am spus că [pe Victor] nimeni nu l-a văzut.

(26)a. Ipe (?to vivlio) **oti** (to vivlio) ton agapai poli (Gr - Anagnostopoulou 1997, 168).

b. Mislja (prestăpnika) če prestăpnika sa go xvanali. (Bg)

- 3) Embedded CLLD Topics typically precede the interrogative complementizers *an/dali* (for lack of space we do not illustrate these cases here).
- 4) Embedded CLLD Topics must precede the subjunctive complementizer/particle *na/da* which, as is well-known, requires strict adjacency with the verb in all of the Balkan languages. In Romanian, the constituent preceding the subjunctive complementizer *să*,

although it can be clitic resumed and therefore may qualify as a CLLD Topic, has to meet the additional requirement of emphasis (Cornilescu 2000), cf.(27).

(27) As dori [**pe Ion**] să-l chemati mîine.

5) CLLD Topics precede the wh-word/phrase in embedded wh-questions, cf. (28):

- (28) a. Dhen ksero afto to vivlio, pjos tha to dhiavasi ja avrio.¹⁰
b. Čudja se tazi roklja koga (li) izobšto šte ja obleka.
c. Mă intreb pe Petre cine-l mai crede.

2. Focus constructions

2.1. Similarities with CLLD Topic constructions

There are a number of distributional similarities between CLLD Topics and focused phrases in the Balkan languages under study. If, as suggested in the literature, the position of Focus is also a result of movement (cf. in particular Tsimpli 1995), the observed similarities can be attributed to the movement nature of Focus phrases, cf. the representations in (4b) above. (29) gives examples of Focus phrases accompanied by a focus particle like *samo/mono*. The focused object phrases can be definite or ‘bare’, i.e. unaccompanied by any definite or indefinite determiner:

- (29) a. *Samo cvetja šte kupja* (ne bonboni); *Samo cvetjata šte i podarja*.
b. *Mono ta luludhia dialeksa moni mou*; *Mono luludhia aghorasa*.

We have observed that focused phrases (in Greek and Bulgarian) can appear: a) displaced in a matrix clause even though they belong to an embedded clause; b) in front of a declarative or an interrogative complementizer; c) in front of a wh-word/phrase in a wh-question.¹¹ All of these properties are attested in Greek and Bulgarian:

- (30) a. [_F **Ti Maria**] lene oti pandreftike o Janis; [_F **ton Jani**] rotisan pjos efighe (Tsimpli 1995, 193).
b. Lene [_F **ti Maria**] oti pandreftike o Janis (Joseph & Philippaki 1987, 104); Mu ipe [_F **to Jani**] oti idhe; Me rotise [_F **ta vivlia**] an aghorasa (Alexiadou 1997, 73).
c. Anarotieme [_F **tu Petro**] ti to edoses; Me rotise [**ta vivlia**] pjos aghorasa.
(31) a. [_F **Maria**] mislja, če šte izberat za predsedatel.
b. Ivan znaex, če šte xodi, no [_F **ti**] če šte xodish, ne znaex (from Rudin 1991)
c. Čudim se [_F **na svekăra**] kakvo da podarim.

Given that these properties are tests for a movement derivation, we can conclude that the dislocated position of Focus is also derived by movement: (30a)/(31a) show instances of unbounded (long-distance) Focus movement; (30b)/(31b) show instances of short Focus movement (to the Left Periphery of a declarative complement c) (30c)/(31c) show instances of short Focus movement in embedded wh-questions.¹²

¹⁰ With certain wh-phrases the CLLD Topic can also be found to the right of the wh-phrase:

(i) a. Dhen ksero pjos, afto to vivlio tha to diavasi ja avrio (Alexiadou 1997, 70)

This seems also true in Bulgarian, although the possibility is attested with ‘heavier’ wh-phrases only.

¹¹ This co-occurrence is not possible in matrix clauses, probably for independent reasons.

¹² This last possibility is also attested in Romanian, according to Cornilescu (2000) who cites cases like (i):

(i) Nu știu **alți** cum sunt, dar **eu** îmi aduc aminte de asta cu plăcere.

Given the data discussed so far, we can generalize that in Greek and in Bulgarian, constituents that can be Topicalized are also eligible for Focalization. In other words, as predicted by the abstract structures in (4) above, the two constructions should be syntactically differentiated through the presence vs. absence of a resumptive clitic (in the case of object noun phrases). This, however, does not seem to be the case in Romanian. As reported by Dobrovie-Sorin (1990), Cornilescu (2000), Alboiu (2000), in this language, not just Topics, but also focused phrases can be clitic resumed:

(32) **Pe Petru** Maria nu l-ar ajuta, pe Gheorghe, da; Eu [_F **pe Popescu**] l-am vazut (nu pe Ionescu); Eu [_F **romanul ăsta**] l-am citit (nu pe calalt). (Dobrovie-Sorin 1990, 220)

While all authors acknowledge that Romanian observes the pan-Balkan ban on doubling of ‘bare’ nouns (i.e. nouns without any determiner), Dobrovie-Sorin (1994) and Cornilescu (2000) nevertheless give examples of focused definite phrases where clitic resumption is not just possible but obligatory even in the presence of focus particles (like *numai* ‘only’, *chiar* ‘even’, *macar* ‘at least’):¹³

- (33) a. **Numai pe Ion** il iubește Maria.
 b. **Macar cartea asta** au citit-o elevii.

The Bulgarian and Greek equivalents of (33) are ungrammatical, as (34) shows:

- (34) a. ***Samo Ivan** go obica Maria (Bg)
 b. ***Mono ton Jani ton** agapai i Maria. (Gr)

As mentioned above, we suggest more generally, that wherever there are differences between the three languages, these seem to be determined by independent language-internal properties. One could think that the contrast between (33) and (34) is a primitive, i.e. non-derived, difference between Romanian and Bulgarian/Greek. But this may well turn out to be related to an independent difference between these languages, namely to the fact that Romanian is not as restricted as Bulgarian and Greek in its use of real clitic doubling (*anticipatio*) where the double is in situ. See the contrast in (35):

- (35) a. L-am văzut numai pe Ion. (Rom)
 b. *Az go vidjax samo Ivan. (Bg)

Whatever the explanation for the distribution of clitic resumed phrases in Romanian focus construction, it is tempting to say that the contrast in (35) is at the basis of that between (33)-(34), if we presume that the clitic doubled noun phrase originates in a postverbal position and then moves to a preverbal position without further changes in the structure. This is rendered plausible by the following two facts:

1) Whenever clitic doubling is impossible in Romanian (as with indefinite quantifiers like *pe altcineva* ‘someone else’ illustrated in (36))¹⁴, resumption of the same phrase in

¹³ According to Cornilescu (2000), in these examples doubling is obligatory because of the inherent semantics of the proper names or of the definite descriptions, which are “good” topics.

¹⁴ According to Alboiu (2000), certain quantifiers (both universal and distributive) like *oricine* ‘anyone’, *fiecare* ‘each’ can be clitic resumed, as opposed to ‘bare’ quantifiers like *fiece* ‘every’, and *cineva* ‘someone’. The author argues that, depending on their inherent semantics, quantifiers behave as CLLD Topics or as Focus. Hence their split behaviour.

preverbal focus is also impossible (which makes one think that *pe* is a necessary but not a sufficient condition for clitic doubling, Dobrovie-Sorin 1994):

- (36) a. *Ion l-așteaptă pe altcineva vs. Ion așteaptă pe altcineva.
 b. *Ion **pe altcineva** l-așteaptă, nu pe Maria vs. Ion **pe altcineva** așteaptă, nu pe Maria

2) Whenever clitic doubling is obligatory in Bulgarian or Greek (as happens with psychological predicates), the fronted focus phrase must also be clitic resumed:

- (37) a. Boli go glavata Ivan (Cf. *Boli glavata Ivan)
 b. (Samo) **Ivan** go boli glavata. (Cf. *(Samo) **Ivan** boli glavata)

2.3. Linear orders of Topic and Focus in the Balkan languages

Finally, another property shared by all of the Balkan languages under study is the relative order of Topics and Focus in the left periphery. In a single clause, there can be multiple CLLD Topics but there is always a single Focus per clause (also known as ‘Focus uniqueness requirement’). Moreover, in conformity to the universal organization of the Left Periphery, Topics must precede all phrases that can be argued to possess a focus feature (Horvath 1986): contrastively focused phrases, bare quantifiers, as well as wh-phrases. There is also a tendency for these latter constituents to appear adjacent to the verbal predicate. Examples are provided below:

- (38) a. [_TMariei] [_Tflorile acestea] tu nu i le poți cumpăra. (Rom -Cornilescu 2001)
 b. [_TMariei] [_F **flori**] este potrivit să-i oferi.
 c. [_T Pe Victor] [_F **cine**]-l așteaptă la aeroport. (Alboiu 2000)
- (39) a. [_TNa Maria] [_T tezi cvetja] sām ì gi podaril az. (Bg)
 b. [_T Marija] [_F **māžāt i**] ja izvika i tja se pribra.
 c. [_T I nego] [_F **koj**] go pita, ama na – kato e za razvala, i toj e tam.
- (40) a. [_TTa vivlia] [_F **sti Maria**] ta edhosa. (Gr - Alexiadou 1997, 74)
 b. Me rotise [_T sti Maria] [_F **pjos**] tis edhose afta ta vivlia.

Based on all of the above comparative data, we can conclude that the overall order of the dislocated phrases in the Greek, Bulgarian and Romanian adheres to the following structural hierarchy:

HTLD > CLLD (CLLD) > FOCUS

3. Conclusion

The organization of the Left Periphery in the Balkan languages, including the relative order of Topic and Focus, reflects a stable typological tendency rather than a pure Sprachbund effect. Nevertheless, the development of the common discourse patterns can be seen as a follow-up process on some of the convergence phenomena (object reduplication and the morpho-syntactic expression of definiteness), which, among other phenomena lead to the establishment of the Balkan Language Union (Assenova 2002). According to Minčeva (1969), Topic structures illustrate some of the most specific properties of the syntax of colloquial speech: shaping of intonational-syntactic groups, the possibility for segmentation of the utterance which “deviates” from the norms of the standard language, ellipsis, pleonasm, etc. These principles have manifested themselves at quite an early stage in the Balkan context. The same could be hypothesized for Focus structures which not only allowed for the

independent syntactic expression of (different kinds of) non-presupposed information, but also create additional stylistic effects. Given the colloquial nature of the bi- and multi-linguistic contacts at the time when the main Balkanisms were integrated into the structure of each language, the universal principles of (colloquial) syntax must have fed the general Balkan tendency towards a greater word order freedom. Topic and Focus are especially relevant for communication purposes, so it is not surprising that such structures have been favoured by speakers in contact situations.

References

- AG (1994) = *Akademična Gramatika na sãvremennija bãlgarski knižoven ezik*. Tom 3. Sintaksis. Sofia, Izdatelstvo na BAN.
- ALBOIU, G. (2000) *The Features of Movement in Romanian*. Doctoral thesis, University of Manitoba
- ALEXIDAOU, A. (1997) *Adverb Placement. A Case study in antisymmetric syntax*. Amsterdam: John Benjamins.
- ANAGNOSTOPOULOU, E. (1994) *Clitic Dependencies in Modern Greek*. PhD dissertation, University of Salzburg.
- ANAGNOSTOPOULOU, E. (1997) "Conditions on Clitic Doubling in Greek". In: van Riemsdijk, H. (ed.). *Clitics in the Languages of Europe*. Berlin: Mouton de Gruyter, pp. 761-798.
- ARNAUDOVA, O. (2001) "Prosodic Movement and Information Focus in Bulgarian", in Franks, S., T. Holloway King & M. Yandroff (eds.) *Annual Workshop on Formal Approaches to Slavic Linguistics. The Bloomington Meeting 2000*. Michigan Slavic Publications, Ann Arbor, pp.19-36.
- ASSENOVA, P. (2002) *Balkansko ezikoznanie. Osnovni problemi na Balkanskija ezikov sãjuz*. Faber.
- CINQUE, G. (1977) "The movement nature of Left Dislocation", *Linguistic Inquiry* 8, pp. 397-412.
- CINQUE, G. (1990) *Types of A' dependencies*. Cambridge, Mass.: MIT Press
- CORNILESCU, A. (2000) "Rhetic Focus at the Left Periphery: The Case of Romanian", presented at the *Going Romance Conference*, Utrecht
- CORNILESCU, A. (2001) *On Focusing and Wh-Movement in Romanian*. Ms., University of Bucharest
- CYXUN, G. (1962) "Mestoimennata klitika I slovoredãt v bãlgarskoto izreãenie", *Bãlgarski ezik*, 1962, XII, 4, pp. 283-291.
- CYXUN, G. (1981) Tipologiãeskie problemy balkanoslavjanskogo jazykovogo areala. Nauka i texnika.
- DOBROVIE-SORIN, C. (1990) "Clitic Doubling, Wh-movement, and Quantification in Romanian", *Linguistic Inquiry*, 21, 3. pp. 351-398.
- DOBROVIE-SORIN, C. (1994) *The Syntax of Romanian*. Mouton de Gruyter.
- DŽONOVA, M. (2004) *Izreãenija sãs semnatiãnata rolja experiercer v sãvremennija bãlgarski ezik*. Doktorska disertacia. Sofia.
- GUENTCHÉVA, Z. (1994) *Thématisation de l'objet en bulgare*. Peter Lang S.A, Bern.
- HORVATH, J. (1986) *Focus in the Theory of Grammar and the syntax of Hungarian*. Fors Dordrecht.
- IVANČEV, S. (1978) *Prinosi v bãlgarskoto i slavjanskoto ezikoznanie*. Sofia: Nauka i izkustvo.

- JOSEPH, B. and I. PHILIPPAKI-WARBURTON (1987) *Modern Greek*. London: Croom Helm.
- KISS, K. É. (1998) "Identificational Focus versus Information Focus." *Language*, 74, pp. 245-273
- KRAPOVA, I. (2002) "On the Left Periphery of the Bulgarian sentence". *University of Venice Working Papers in Linguistics*, vol. 12, 2002
- LINDSTEDT, J. (2000) "Linguistic Balkanization: Contact-induced change by mutual reinforcement" In: Gilbers, D., J. Nerbonne, & J. Schaeken, (eds.) *Languages in Contact*. Amsterdam & Atlanta, GA: Rodopi. (Studies in Slavic and General Linguistics, 28.), pp. 231-46
- LOPAŠOV, Ju.A. (1978) *Mestoimennye povtory dopolenija v balkanskix jazykax*. Leningrad.
- MINČEVA, A. (1969) "Opit za interpretacia na modela na udvoenite dopälnenija v nälgarskija ezik", *Izvestija na Instituta za bälgarski ezik*, XVII, Sofia, pp. 3-50.
- MLADENOV, M. St. (1965) "Beležki värxu sintaksisa na ixtimanskija govor", *Bälgarski ezik*, 1965, 3, pp. 208-224.
- PHILIPPAKI-WARBURTON, I., S. Varlakosta, M. Georgiafenius, G. Kotzoglou. "Moving from theta-positions: pronominal clitic doubling in Greek", *Lingua*, pp. 963-989.
- POPOV, K. "O logičeskom udarenii", *Voprosy jazykoznanija* 1961, 3.
- POPOV, K. "Stilno-gramatična upotreba na udvoenoto dopälnenie v bälgarskija knižoven ezik". *Izvestija na Instituta za bälgarski ezik*, XVIII, Sofia, pp. 459-470.
- STOYKOV, S. (1962/2002) *Bälgarska dialektologija*. Sofia: Akademično izdatelstvo "Prof. Marin Drinov".
- TSIMPLI, J.-M. (1995) "Focusing in Modern Greek". In: K. Kiss (ed.) *Discourse-Configurational Languages*. Oxford: Oxford University Press.
- RIZZI, L. (1997) "The Fine Structure of the Left Periphery", in L. Haegeman, (ed.) *Elements of grammar*, Dordrecht: Kluwer, pp.281-337
- RUDIN, C. (1986) *Aspects of Bulgarian Syntax: Complementizers and wh-constructions*. Slavica Publishers, Inc., Columbus, Ohio.
- RUDIN, C. (1994) "Topic and Focus in Bulgarian". *Acta Linguistica Academiae Scientiarum Hungaricae*, vol. 40 (3-4), pp. 429-447.
- VALLDUVÍ, E.(1992) *The Informational Component*, New York, Garland
- ZUBIZARRETA, M.-L.(1998) *Prosody, Focus and Word Order*. MIT Press, Cambridge, Mass.