

## *On the syntagmatic typology of Modern Bulgarian*

Valentin Geshev

Sofia University St. Kliment Ohridski

В статье вводится понятие *синтагматическая типология*, которое отражает взаимодействие разных синтагматических механизмов, таких как порядка слов, согласования, формальных падежей, лексического значения, одушевленности, формальных показателей частей речи, просодии, определенности, удвоения дополнения, экстралингвистических знаний. На этой основе выявляется синтагматический тип *согласовательных индоевропейских языков в Европе*. Современный болгарский язык отнесен к беспадежному подтипу с учетом важной роли согласования при сохранении сравнительно свободного коммуникативно значимого порядка слов после радикальной редукции его падежной системы.

The paper frames the concept of *syntagmatic typology* based on the interaction among various syntagmatic mechanisms, such as word order, congruence, formal cases, lexical meaning, animacy, formal class markers, prosody, definiteness, pronominal reduplication of the object and extralinguistic knowledge. On these grounds the syntagmatic type of *congruence Indo-European languages in Europe* is abstracted. Modern Bulgarian is classified as a model representative of its case-less subtype because of the extraordinary role congruence plays in the preservation of its relatively free word order, specialized to express the theme-rheme structure of the sentence, after a substantial reduction of its case system.

*Key words:* syntagmatic typology, syntagmatic mechanisms, congruence languages, Modern Bulgarian, theme-rheme structure, lack of formal cases.

### **1. Heuristic introduction.**

While working on the history of formal cases in Bulgarian I came to the intriguing question of how Modern Bulgarian succeeds in combining its relatively free word order, inherited from earlier highly inflectional linguistic stages, with a complete lack of formal cases with nouns, adjectives and some classes of pronouns. Looking for the answer to this question I concentrated on the interaction between word order, congruence<sup>1</sup>, formal cases, lexical

---

<sup>1</sup> Used synonymously to *agreement*. In English *congruence* is a rare and somewhat old-fashioned linguistic term. Thus, it is regularly used by Leonard Bloomfield (see 5.) but not by Otto

meaning, formal class markers, prosody, etc. as means to explicate the concrete syntactic relations in a sentence. In the present paper these means will be called *syntagmatic mechanisms*.

In Geshev (2013) (and two other forthcoming papers) I came to the conclusion that in Bulgarian the most important mechanism to compensate for the lack of formal cases is congruence, which is often, but not always, able to eliminate potential ambiguity in the sentence, assisted to a certain extent by animacy, definiteness, pronominal reduplication of objects and extra-linguistic knowledge. On this ground I proposed the notion of *congruence Indo-European languages in Europe* – a typological grouping which includes a substantial part of Europe’s languages – and I classified Modern Bulgarian as belonging to the case-less subtype of these languages. All this should provoke further speculation on the importance of the concrete syntagmatic mechanisms for the disambiguation of the sentence and on the typological significance of their interaction, which is the topic of this paper.

## 2. Syntagmatic mechanisms.

Generally speaking, all mechanisms which connect syntagmatically – formally or semantically – two or more linguistic units might be called syntagmatic mechanisms, but, if we confine ourselves to the syntactic level, by syntagmatic mechanisms we may mean any mechanisms connecting syntagmatically two or more words in a sentence, and this is the sense in which I shall use the expression *syntagmatic mechanisms* in the present paper. Defined in this way it designates a notion which covers several linguistic phenomena of utmost importance to syntax and typology.

Syntagmatic mechanisms are easily recognizable if we pose the question of how the users of a language<sup>2</sup> decide which words or phrases are immediately connected in a sentence or in a text.

**2.1.** The mechanisms may be **semantic** and, thus, dependent on our extralinguistic knowledge. If a person with a low competence in a foreign language has succeeded in discerning only the words *wolf*, *goat* and *grass* in a sentence or in a short text, it is more than possible that she or he would imagine a typical scene of everyday life of food chains. This might prove to be a false interpretation of what has been heard or read, but it would be the most natural

---

Jespersen (1969) or John Lyons (1968), who prefers *concord*. Furthermore, Lyons uses congruence as philosophic but not as grammatical term. The current page on *agreement* in Wikipedia gives only *concord* as its synonym. My preference for this neo-Latinism is based on its being in harmony with Central European linguistic tradition where it is widely used usually as synonymous to native terms preferred by purists, cf. German *Kongruenz* and *Übereinstimmung*, Czech *kongruence* and *shoda*, Polish *kongruencja* and *zgoda*, Croatian *kongruencija* and *sročnost*.

<sup>2</sup> The users of a language as a set larger than its native speakers.

semantic relation between the three words based on the typical interaction of wolves, goats and grass in life. This agreement in meaning, analogous to formal agreement, is clearly manifested in pairs of words like *graze* and *grass*, *bird* and *fly*, *swim* and *water*, *plough* and *land*. Linguists have been aware of this for a long time and they have given various names or explanations for it<sup>3</sup>. What is common to **semantic and formal agreement** is the repetition of certain meaning in two syntagmatically connected linguistic units<sup>4</sup>, i.e. the role of **redundancy** as an instrument of syntagmatic cohesion and as a sign of the immediate syntagmatic relation between the respective units. This should be a role more or less different from the role of redundancy in information theory<sup>5</sup>. Redundancy puts the sentence or the text together in the way bricks overlap to build up the wall.

**2.2. Linear contact** often marks the immediate syntagmatic relation between two linguistic units. In most instances it is obligatory for the correct coding and decoding of the sentence. If we dis sever the adjectives (or other kinds of attributes) or the adverbs from the words they modify – even in languages with well developed formal cases and congruence – and place them somewhere else in the sentence, even relatively short syntagms become almost unintelligible and are perceived as **formally inappropriate**. The same applies to the dislocation of prepositions and conjunctions. Thus, if we dislocate the words of a Latin (1) or a Ukrainian (3) syntagm at random, we may obtain a certain number of thoroughly unintelligible sequences of forms used inappropriately as in (2) and (4):

(1) In cornu tauri parva sedebat musca.  
on horn of-bull tiny was-sitting fly<sup>6</sup>

(2) → tauri sedebat parva cornu in musca  
of-bull was-sitting tiny horn on fly

(3) I blidyj misjac' na tu poru Iz xmary de-de vyhljadav... (T. Shevchenko)  
and pale moon at that time from cloud hardly peeped

(4) → vyhljadav iz tu misjac' i poru blidyj na de-de xmary  
peeped from that moon and time pale at hardly cloud

---

<sup>3</sup> Thus, for instance, in 1938 Aleksandar Belić speaks about the unity of meaning of the syntagm. In his opinion the subordinate member of a syntagm reveals (or, I would say, introduces) a quality, an object, an instrument, etc., latently present in the notion represented by the head of the syntagm (Belić 1938).

<sup>4</sup> Vladimir Skalička claims that the repetition of a part of the meaning is inherent to any combination of lexical or morphological units: "... every word, every morpheme already contains a part of the next word, of the next morpheme. This also means that every word, every morpheme repeats a part of the previous word or morpheme. It repeats something and adds something new, at least the repetition itself." (Skalička 2004: 168).

<sup>5</sup> Cf. Geshev (1992: 38–39).

<sup>6</sup> Only the categories relevant to the topic discussed are glossed, in this example only word order.

The word-for-word English translation of the syntagms when disarranged gives only a vague idea of their formal inappropriateness – it illustrates the wrong place of the words in them but not the unacceptability of the case and gender forms in the linear combinations they have got into.

**2.3.** In natural languages there is a strict and multistage hierarchy of the parts of the sentence which has possibly evolved for hundreds of thousands or even millions of years. Various linguistic schools visualize this hierarchy by means of tree diagrams, combinations of brackets, etc. If the user of a language is able to discern the parts of a sentence, the syntagmatic relations (immediate or remote) between the concrete linguistic forms in it will be clear enough to her or him. So the **formal characteristics of the parts of the sentence** (concerning their morphology, word order and lexical classes) are an important means to mark their syntagmatic relations, i.e. they are a syntagmatic mechanism. Practically, we often rely on **discerning the parts of speech** (the word-classes) with their salient categories and their predominant usage as concrete parts of the sentence. Thus, if we identify the finite verb in a clause it should be its predicate, and if we find a noun in the nominative it should be its syntactic subject in the majority of cases, so both items would be immediately connected syntagmatically.

### **3. Syntagmatic typology.**

The interaction between the different syntagmatic mechanisms is language-specific. It is an important typological characteristic on the basis of which languages can be grouped and classified. The notion of syntagmatic mechanisms includes formal and semantic techniques operating on various – morphological, lexical, syntactic, semantic, cognitive – levels. As the expression *typology of the interaction between syntagmatic mechanisms* is too clumsy I would prefer to reduce it to *syntagmatic typology* and this is the tentative term I suggest for the phenomenon under consideration. So by **syntagmatic typology** I mean a typological characteristic based on the specific combination and interaction of the syntagmatic mechanisms in a given language or group of languages and by **syntagmatic type** – the concrete manifestations of syntagmatic typology.

Being concrete combinations of syntagmatic mechanisms, the syntagmatic linguistic types are most probably limited regionally and genetically.

Indeed, although the principle task of typology is to analyze (qualitatively, quantitatively and ontologically) the differences and similarities between the languages throughout the world<sup>7</sup>, when typologists abstract a concrete linguistic type it often turns out to be limited regionally and genetically. Within the framework of content typology, Georgij Klimov distinguishes the so-called class languages as one of the major stages in the evolution of linguistic

---

<sup>7</sup> For a more detailed approach cf. Bickel (2005).

structure. However, most representatives of this type are closely related – they belong to the Bantu languages – and are confined to Central and South Africa. Another linguistic type, whose originality and distinctive features have been established for the first time by G. Klimov, is the active one – although the genetic relations among its modern representatives are not so clear, they are to be met mainly in the Americas<sup>8</sup>.

The Philippine morphosyntactic alignment – one of the several determined on the basis of universal criteria in a world-wide classification – is characteristic of languages genetically related and geographically limited to the Philippines, Borneo, Taiwan and Madagascar. Martin Haspelmath's European linguistic area is a further example of a typological grouping determined geographically and genetically. It is significant that M. Haspelmath tries to find the reasons for the appearance of various isoglosses in this area, relying on the history of the concrete Indo-European languages in Europe (Haspelmath 2001: 1506–1507).

#### 4. The syntagmatic type of Modern Bulgarian.

Linguists in Bulgaria traditionally say that Modern Bulgarian is an analytic language. This characterization is inherently superficial and inexact. The Modern Bulgarian noun is case-less but the categories of gender, number and definiteness are expressed synthetically, and with the verb all finite forms are synthetic or semi-synthetic<sup>9</sup>.

**4.1.** Two of the features of Modern Bulgarian are of utmost importance for syntagmatic typology – the **lack of formal cases** with nouns, adjectives and some classes of pronouns and the **relatively free word order**, which is able to express the theme-rheme structure of the sentence more or less independently of its division into formal syntactic categories such as subject, object, etc. Standard Bulgarian has been completely deprived of cases with nouns since the middle of the twentieth century. Written Bulgarian still distinguishes between a nominative and an oblique case with definite masculine nouns and their attributes, but the phonetic distinction between these forms is negligible and in colloquial Bulgarian they are used interchangeably with no case-like function. So, when we consider the syntagmatic typology of languages like Modern Bulgarian, the main question is why and how the inherited free word order has been preserved after the loss of formal cases with nouns and their attributes.

**4.2.** In this respect Bulgarian is not alone – it belongs to a major subtype of a syntagmatic type encompassing a **great part of the languages in Europe**, remarkable for the **extraordinary role of congruence** in the linguistic

---

<sup>8</sup> Cf. e.g. Klimov (1977); in English Klimov (1974). G. Klimov himself uses the terms *kontensivnyj* in Russian and *contentive* in English. Further investigations establish active typology features outside the Americas (cf. Andréasson 2001) and class typology features outside Africa (cf. Aikhenvald 2000).

<sup>9</sup> Cf. Geshev (1990).

structure as a whole, and for the preservation of a relatively free word order after a substantial reduction of the case system in particular. These languages are characterized by:

- a) well-developed agreement between the subject and the predicate;
- b) well-developed agreement between nouns and their attributes;
- c) relatively free word order;
- d) presence or absence of formal cases;
- e) Indo-European origin;
- f) being indigenous to Europe.

This combination of features may serve as a tentative definition of a type which we may call **congruence Indo-European languages in Europe**<sup>10</sup>.

In relation to feature (d) Modern Bulgarian belongs to the **case-less subtype** of the congruence Indo-European languages in Europe. Being a model representative of this subtype, it reconciles free word order with the lack of formal cases through a specific interaction between its syntagmatic mechanisms.

## 5. The importance of congruence.

By congruence I mean any kind of formal agreement between two (or more) items in a sentence and especially between the noun and its attributes and the subject and its predicate.

Of course, other approaches are also possible. Leonard Bloomfield defines the agreement in the “nominative expression”, i.e. between nouns and their attributes, in the majority of the Indo-European languages as well as the agreement in the “actor-action construction”, i.e. between the subject and the predicate, in Modern English as “concord” or “congruence”, while the agreement in the Latin expression *puella cantat* ‘(the) girl she-sings’ is defined by him as “cross-reference” (Bloomfield 1970: 191–193). Such a subtle differentiation would be irrelevant to our purposes. For the same reason, I would regard the role of formal cases as essentially different from congruence, though both phenomena may be grouped together on certain grounds, as it is, for instance, in Christian Lehmann’s classification distinguishing between “unirelational and birelational grammatical formatives”: “The most important birelational formatives are pronominal elements marking cross-reference and adverbial/adpositional elements marking case relations.” (Lehmann 2002: 154).

**5.1.** In the languages considered, congruence plays a key role in the **noun phrase**. It should be noticed, however, that there is a wide variety of noun phrase morphological and word order patterns within this group.

---

<sup>10</sup> Cf. Geshev (2013: 378). The question of which languages belong to this type is discussed in Section 6.

If in Modern Bulgarian or in Modern Greek the adjective usually precedes the noun, e.g. Bulg. *hubava nošt*, Gr. *kalí nýxta* ‘good night’, in Albanian it obligatorily comes after the noun and, what is more, there is an obligatory copulative article between the noun and the majority of adjectives, e.g. *natë e mirë* ‘night which good’, i.e. ‘good night’. We can observe a similar variety with the grammatical categories manifested in the noun phrase. Besides definiteness of nouns, which is grammaticalized in some languages but not in others, we may mention adjectival categories non-existent in Modern Bulgarian – cf. the contrast between prepositive and postpositive adjectives in Spanish (*buen amigo* against *amigo bueno* ‘good friend’) or the opposition of simple to extended forms of the attributive adjectives in Lithuanian (*pirma pamoka* against *pirmoji pamoka* ‘first lesson/lesson one’; the same opposition is inherent to Old Bulgarian and, residually, to Serbian, Croatian and Slovenian). However, it is mainly characteristic feature (a) in the tentative definition proposed above – the **well-developed agreement between subject and predicate** – which is of primary concern to our topic, especially if considered as a means to compensate for the lack of cases in a language where the communicative function of a relatively free word order is preserved. This is determined by our chief interest in the syntagmatic mechanisms immediately connecting the predicate in a clause with its syntactic subject and direct object.

5.2. In Bulgarian, where subjects and direct objects are equally expressed by noun phrases without prepositions, the **agreement between subject and predicate** is the only formal means to unambiguously identify these syntactic roles, and it happens to be the only formal marker of the subject when there are no other formal or semantic differences between subject and object noun phrases, cf. sentences (5) and (6) in which both nouns (except for the circumstantial phrase) are equally indefinite and animate:

(5) Dete v džunglata otkriha vojnici.  
 child in-the-jungle found(pl) soldiers  
 ‘A child was found in the jungle by soldiers.’

(6) Vojnici v džunglata otkri dete.  
 soldiers in-the-jungle found(sg) child  
 ‘Soldiers were found in the jungle by a child.’

Here the noun phrases *dete* ‘a child’ and *vojnici* ‘soldiers’ differ in number and there is no ambiguity about their being a subject and a direct object, since it is the subject which agrees in number with the verb. So in (5) the plural form *vojnici* is the subject, in spite of its sentence-final position, because the verb *otkriha* is also in the plural; analogically in (6) the sentence-final noun *dete* is the subject because it agrees in number with the singular verbal form *otkri*. If

the intonation of both sentences is not emphatic or inverse<sup>11</sup> they would have sentence-final rhematic subjects and sentence initial thematic direct objects<sup>12</sup>.

With a lot of Bulgarian verbal categories the subject and the predicate agree also in gender and then the different gender of the noun phrases may also help. Thus, congruence is an unequivocal instrument of pointing to the subject of a sentence, but it cannot work when the preposition-less noun phrases **share the same noun-class characteristics**, i.e. when they are of the same gender and in the same number.

**5.3.** Congruence tends to be underestimated as a **plausible reason for the preservation of a relatively free word order** after the loss of formal cases. This could be illustrated by the speculations of P. Sgall, E. Hajičová and E. Buráňová (Sgall et al. 1980: 142–149) about the typological characteristics of functional sentence perspective (called “topic/comment articulation” in the English summary of their book – *ibid.*: 154). The authors note that, concerning word order, there are significant distinctions between the languages in Western Europe characterized by isolating (analytic) typological features – in English and French functional sentence perspective is marked mainly by intonation and grammatical constructions, while in Spanish and German word order is relatively free (and closer to that in the Slavonic languages) and it serves to express functional sentence perspective. The only reason for such a freedom of word order, mentioned by the authors, is the capability of Spanish to mark the animate direct object through the preposition *a* (Sgall et al. 1980: 146–147), i.e. free word order is explained through the existence of a construction similar to formal cases and not through the presence of congruence. The chain of causation is rather the opposite: in medieval Spanish the construction with the preposition *a* began to mark the animate direct object because there was a relatively free word order inherited, and it was not free word order which appeared because a special construction to mark the direct object already existed. There is a similar construction – with the preposition *pe* – in Rumanian, too.

So a stage with no cases and free word order is attested in medieval Spanish and Modern Bulgarian. This refutes O. Jespersen’s claim that fixed word order is the cause and “abolition of case distinction” the effect (Jespersen 1969: 361).

## **6. Representatives of the type.**

The strength of congruence as a syntagmatic mechanism varies throughout languages, so it is difficult to distinguish between congruence and “non-congruence” languages in a thoroughly precise manner.

---

<sup>11</sup> By inverse intonations I mean intonation which marks the inverse rheme-theme linear arrangement of a sentence, i.e. falling intonation on its first segment and even low intonation on its second fragment.

<sup>12</sup> Theme and rheme as defined in Section 7.1.



If in Europe we move from South to North and from East to West the role of congruence as a syntagmatic mechanism decreases because of the:

- a) bleaching or loss of gender (almost complete in English);
- b) bleaching or loss of the personal endings of the verbs (complete in the major Scandinavian languages).

More or less typical congruence languages are all Slavonic and all Indo-European Balkan ones. Some of them possess well-developed case systems, others (like Rumanian) use only a few formal cases, and some of them use no cases with nouns. The last subtype is represented by Modern Bulgarian and the West Romance languages, including Occitan in Southern France but not French proper.

Continental Germanic (the German-Dutch linguistic continuum) may also be classified at least as a relative representative of the type – in spite of the obligatory use of subjects and some specific word order rules in it. This means that not all congruence Indo-European languages in Europe are null-subject (sub-drop) languages, although the majority of them are. This null-subject majority is what M. Haspelmath calls *referential-agreement languages* – he specifies **Bulgarian** as their model representative and points out that they are far more widespread than the languages with an obligatory use of overt subjects and concomitant agreement between subjects and predicates. He calls the latter *strict-agreement languages*; and further notes that they are quite rare but some of them – English, French and German – “happen to be well-known” (Haspelmath 2001: 1500).

With the fixed word-order position of their syntactic subjects and direct objects, English and French cannot represent the syntagmatic type considered, though morphological agreement is not completely alien to them. The Scandinavian languages should be rather considered as not belonging to the type, in spite of their employing gender in the noun phrase, because they lack personal endings with the verb (this does not concern the more archaic Icelandic and Elfdalian).

There is a certain weakening of the function of personal endings also in languages which should still be classified as congruence ones. The East Slavonic verb has no personal endings in the past tense and in the conditional (but these forms agree with the subject in gender and number)<sup>13</sup>. In Lithuanian and Latvian, and with certain verbs also in Rumanian and Czech, third person verbs do not distinguish between singular and plural forms – a distinction which is very important for the disambiguation of the Bulgarian sentence, as we saw in examples (5) and (6). The well-developed case systems of Lithuanian, Latvian, East Slavonic and Czech succeed in compensating for this “shortcoming”, while Rumanian relies on the preposition *pe*, which introduces animate direct objects, and on the employment of pronoun reduplication of objects.

---

<sup>13</sup> This makes the use of overt subjects in Russian almost obligatory – cf. A. Kibrik (2013).

## 7. Other interacting syntagmatic mechanisms.

Except for congruence, there are several other syntagmatic mechanisms participating in coding and decoding the syntactic structure of the Bulgarian sentence, such as:

- a) definiteness;
- b) animacy;
- c) pronominal reduplication of objects;
- d) extralinguistic knowledge;
- e) context;
- f) intonation.

Word order, though a syntagmatic mechanism in itself, is not included in the list because it serves rather to mark the theme-rheme structure of the sentence (its functional sentence perspective) than to distinguish between subjects and direct objects. Nevertheless, word order is of great importance for our topic and, as a rule, Bulgarian word order generates a serious interest among linguists, usually in combination with other linguistic phenomena – cf. Jordan Penčev (1980), Donald Dyer (1992), John Leafgren (2002), for instance.

**7.1** I stick to the opinion that the **theme-rheme structure** of a sentence is different from its articulation into given and new information<sup>14</sup>. The simple definition, proposed by C. Lehmann (2002: 95), that “theme and rheme” are “that about which something is said, and what is said about it” is good enough for our purposes<sup>15</sup>. What is more, I consider the distinction the same author makes between theme-rheme and topic-focus extremely useful:

... topic and focus, as they appear in left-dislocation and clefting, are completely free and wild, as it were, since they transcend the bounds of the simple sentence; whilst theme and rheme may be considered as tamed forms of the topic and the focus, respectively, since they may structure the simple sentence. In a parallel fashion, the intonation contour is narrowed down on the way from topic/focus to theme/rheme: the pause after the topic, and the contrastive stress on the focus, are reduced. (Lehmann 2002: 105–106)

So in this paper I shall analyze only non-emphatic sentences with “tamed” themes and rhemes, often with rhematic subjects, conscious of the latter being exotic to speakers of English. To simplify things, sentences with inverse rheme-theme linear arrangement and corresponding intonation contour are almost excluded from the present analysis. Passive sentences are also neglected.

---

<sup>14</sup> For a detailed discussion on this problem cf. A. Bogusławski (1977: 147–155)

<sup>15</sup> Bogusławski’s (1977: 142) definition is practically the same: “thing spoken about and what is said about it”.

**7.2. Definiteness** and **animacy** seem to be helpful in telling the subject from the direct object when, say, both preposition-less noun phrases are in the same number and of the same gender and congruence is unable to identify the subject. Thus in (7) the sentence-final noun *pomoštničkata* ‘the assistant’ will be the rhematic subject because it is definite, unlike the sentence-initial phrase *dobra sekretarka* ‘a good secretary’, which is indefinite, and hence the thematic direct object; both noun phrases are animate:

(7) Dobra sekretarka šte nameri pomoštničkata mi.  
 good(f, sg) secretary(f, sg) will find(sg) the-assistant(f, sg) my  
 ‘A good secretary will be found by my assistant.’

Respectively, in (8) *učenik* ‘a pupil’ will be the rhematic (post-verbal) subject and *stranen kamāk* ‘a strange stone’ the thematic (pre-verbal) direct object because *učenik* is animate and *stranen kamāk* is not:

(8) Stranen kamāk otkri učenik izvān grada.  
 strange(m, sg) stone(m, sg) found(sg) pupil(m, sg) out-of-the-town  
 ‘A strange stone was found by a pupil out of the town.’

**7.2.a.** However, definite and animate noun phrases are **preferable subjects only statistically** – because of the common elements in the semantic (including pragmatic) motivation of the syntactic subject, morphological definiteness and lexical animacy. Even if we are interested only in non-emphatic sentences containing only “tamed” themes and rhemes – considering, after C. Lehmann (2002: 100–107), the topic and the focus as emphatic forms of the theme and the rheme – we can find counter-examples in which the animate and definite noun is the (thematic, pre-verbal) direct object and the inanimate and indefinite one is the (rhematic, post-verbal) subject, as in (9):

(9) Gradčeto razvālnuva neočakvano sābitie.  
 the-town(n, sg) excited(sg) unexpected(n, sg) event(n, sg)  
 ‘The town was excited by an unexpected event’

Here *gradčeto* ‘the town’ should be the statistically preferable subject because of its definiteness and (quasi-)animacy (it stands for the inhabitants of the town) and because of its being the sentence-initial theme (themes are also preferable subjects<sup>16</sup>). But the **lexical meaning** (and the intention) of the “affective” verb *razvālnuva* ‘excited’ supposes an animate direct object,

---

<sup>16</sup> Independent of the fact that subjects may be preferable themes. As for Bulgarian, the latter should be proved statistically.

so in reality the inanimate, indefinite and sentence-final phrase *neočakvano sǎbitie* ‘an unexpected event’ is the subject and the quasi-animate, definite and sentence-initial noun *gradčeto* is the direct object in this sentence.

**7.2.b.** A merger of **lexical semantics and extralinguistic knowledge** specifies the subjects in (10):

- (10) Stenata udari snarjad, a kǎštata raketa.  
the-wall hit(sg) shell and the-house rocket  
‘The wall was hit by a shell and the house by a rocket.’

This sentence consists of two clauses in which four equally inanimate noun phrases, all of them in the singular, share a common predicate, also in the singular. On morphological and linear arrangement grounds the definite and thematic *stenata* ‘the wall’ and *kǎštata* ‘the house’ should be the preferable subjects in the sentence, but they are just the direct objects in it, while the indefinite and rhematic *snarjad* ‘shell’ and *raketa* ‘rocket’ are the subjects for the simple reason that we know that shells and rockets can hit walls and houses and not vice versa.

**7.3. Pronominal reduplication of objects**<sup>17</sup> is not reliable enough as an instrument for disambiguating the formal structure of a sentence for several reasons.

**7.3.a.** First, **prescriptive stylistics** in Bulgaria – as far as it exists in some form or another – discourages this syntactic technique. Usually, institutionalized grammars describe it as inherent to colloquial speech and even to dialects<sup>18</sup>, so teachers, proof-readers and editors are overzealous in not allowing it into written texts. As a result, practically no Bulgarian was taught to use it at school and she or he would readily omit a pronominal clitic which reduplicates an object in pursuit of a more bookish (or more “refined”) style.

**7.3.b.** Second, the pronominal clitic **always stands immediately before the verb** no matter whether the object it “doubles” precedes or follows the verb. So we could equally well say:

- (11) Dǎrvetata gi izseče brat mi.  
the-trees them cut(sg) brother my  
‘The trees were cut by my brother.’

and

- (12) Brat mi gi izseče dǎrvetata.  
brother my them cut the-trees  
‘My brother cut the trees.’

---

<sup>17</sup> On pronominal clitics in Bulgarian cf. Tania Avgustinova (1994), Ludmila Uhlřıfova (2011).

<sup>18</sup> As, for instance, K. Popov and E. Georgieva do in *Sintaksis* (1983: 186–188, 282 – 283).

In both sentences the syntactic subject is *brat mi* ‘my brother’ and the direct object is *dārvetata* ‘the trees’, but the direct object is in sentence-initial, pre-verbal position in (11) and in sentence-final, post-verbal position in (12). Notwithstanding the different linear position of the object, it is reduplicated by the pronominal clitic *gi* ‘them’ which is always pre-verbal. There is no ambiguity about the syntactic structure of (11) and (12) because congruence marks the only singular noun phrase in them as their subject and the pronominal clitic reduplicating the object marks the only plural noun phrase as their direct object.

**7.3.c.** However, simple sentences may be **morphologically and syntactically ambiguous** in Bulgarian if they contain preposition-less noun phrases which do not differ in number, gender, and definiteness: then the verb could agree with and the pronominal clitic could double either noun phrase in them:

- (13) *Kāštata ja prodade bankata.*  
 the-house(f, sg) it(f, sg) sold(sg) the-bank(f, sg)  
 (most probably) ‘The house was sold by the bank.’
- (14) *Bankata ja prodade kāštata.*  
 the-bank(f, sg) it(f, sg) sold(sg) the-house(f, sg)  
 (most probably) ‘The bank sold the house.’

In (13) and (14) the verb *prodade* ‘sold(sg)’ and the pronominal clitic *ja* ‘it(f, sg)’ might equally well agree with either *kāštata* ‘the house’ or *bankata* ‘the bank’. The only reason to decide that in both sentences *the bank* is the most probable subject is our knowledge that banks usually sell houses. Here is the third reason for the unreliability of the pronominal reduplication of a direct object as its marker – this mechanism cannot work if there are **no categorial distinctions** between the preposition-less noun phrases in the clause. Under the same conditions the subject-predicate agreement is also unreliable.

**7.3.d.** Sentences (12) and (14) are possible with **falling intonation** on the initial noun phrase as if they answer the questions *Who cut the trees?* and *Who sold the house?* Then (the Bulgarian words for) *my brother* and *the bank* will be sentence-initial but rhematic subjects and the sentences will have an inverse (rheme-theme) communicative structure. Frequently enough sentences like (12) and (14) are to be uttered with falling intonation on the verb as if they answer questions like *What did your brother do with the trees?* and *What did the bank do with the house?* Then (the Bulgarian for) the verbs *cut* and *sold* will be the rheme, while the pre-verbal subjects *my brother* and *the bank* as well as the post-verbal objects *the trees* and *the house* will be thematic. Such sentences with “split themes” are investigated by Yovka Tisheva (2014: 56–60) on colloquial Bulgarian corpus material. It is precisely split themes which could prove her

and Marina Dzhonova's claim that pronominal reduplication of objects is an instrument of topicalization in Bulgarian (Tisheva, Dzhonova 2006: 236), because thematic objects are "doubled" even in post-verbal position<sup>19</sup>.

If (12) and (14) are possible with falling intonation on the final noun – as if answering the questions *What did your brother do?*, *What did your brother cut?* or *What did the bank do?*, *What did the bank sell?* – (the Bulgarian for) *the trees* and *the house* will be sentence-final post-verbal rhematic objects doubled by a pronominal clitic and this will be in unison with Svetomir Ivančev's (1978: 128–152) claim that **rhematic objects may be optionally doubled** in colloquial Bulgarian. However, this statement is controversial and needs further study. Were it real, it would be a fourth reason to consider reduplication of objects but a relatively reliable means to disambiguate the sentence.

**7.3.e.** The fifth reason for the unreliability of pronominal reduplication of direct objects is that in colloquial Bulgarian **direct objects which are not definite or individualized** are usually not doubled by a pronominal clitic. Thus inherently ambiguous sentences can immerge and the usage of equivocal titles (carrying only new information) in the Bulgarian press, such as (15) and (16), is an ample proof that native speakers of Bulgarian are conscious of the ambiguity their language allows for:

(15) Kon krade ciganin

horse steals Gipsy

(16) Mečka ubi germanec

bear killed German

In these sentences two animate and indefinite nouns in the singular surround symmetrically a verb which is also in the singular – a typical case of inherently equivocal sentences, consciously searched for by journalists for the purpose of irony. The only key to their disambiguation is our **extralinguistic knowledge**. Just because we know that horses usually do not steal and that foreigners come to Bulgaria on hunting tourism we hope that it was the Gipsy who stole the horse and the German who shot down the bear, although the opposite might also be expected.

## **8. Typological parallels.**

If, from a certain, not so rigorously scientific, point of view, we qualify the interaction of word order with other syntagmatic mechanisms in Bulgarian as picturesque or exotic, we may speak of a local, or rather areal **typological and word order colouring** (or simply specific combination of features), with

---

<sup>19</sup> As for the possibility that a clause may have a multiple and gradual theme-rheme structure cf. Sgall et al. (1973) and A. Bogusławski (1977).

parallels in the rest of the congruence Indo-European languages in Europe, especially in those of them which lack cases with the noun. Sometimes (but not always!) the parallelism is striking, cf. the Spanish-Bulgarian syntactic, linear and categorial isomorphism in examples (17) and (18), taken from Milena Popova (2012: 144, 149):

(17) Al cazador le disparó el arma.  
to-the hunter him shot the gun  
Na loveca mu grāmna puškata.  
to the-hunter him shot the-gun  
'The hunter's gun went off / The hunter shot his gun.'

(18) Eso me lo dijiste ayer.  
this me it you-said yesterday  
Tova mi go kaza včera.  
this to-me it you-said yesterday  
'You told me this yesterday / This is what you told me yesterday.'

But a feature really common to the congruence Indo-European languages in Europe is their ability to place the verb before the subject and the direct object before the verb, cf. examples (19), (20) and (21), taken from Spanish, Greek and Old Provençal poetry:

(19) Por el agua de Granada sólo reman los suspiros. (F. G. Lorca)  
on the water of Granada only row(pl) the sighs  
'Only the sighs row down the water of Granada.'

(20) Típot' állo dhe vgházun tis tyránnias i klóni. (Pieridis)<sup>20</sup>  
nothing else not bear(pl) of-the of-tyranny the branches  
'No other fruit has been born by the branches of tyranny.'

(21) E platz mi, quan li corredor Fan la gens... (B. de Born)  
and pleases me when the raid make(pl) the people  
'And I am delighted at the people's carrying out a raid...'

Sentences (19), (20) and (21) have sentence-final post-verbal rhematic subjects – *los suspiros* 'the sighs', *i klóni* 'the branches', and *la gens* 'the people' – and (20), (21) have pre-verbal direct objects – *típot' állo* 'nothing else', *li corredor* 'the raid'; (19) is intransitive. If translated into Bulgarian all these sentences may retain (almost) the same word order without losing their grammaticality or the semantic (including pragmatic) peculiarities of the original.

<sup>20</sup> Quotation taken from Papageorgiou (1985: 92)

## 9. Useful and convenient instrument.

If we try to answer the question of **why there is no tendency to transform the theme of the sentence into its (usually pre-verbal) syntactic subject** after the loss of nominal formal cases in Bulgarian, as it has happened in English<sup>21</sup> (provided that in Winfred Lehmann's (1976) opinion this is an ancient tendency in the Indo-European languages), we may suppose that:

First, completely equivocal sentences (that cannot be disambiguated either by our extralinguistic knowledge or by the specific context) are **statistically negligible**.

Second, the speaker can avoid ambiguity by **changing her or his communicative strategy**, using, say, emphatic or passive constructions, unequivocally distinguishing between the parts of the sentence.

Third, relatively free word order, more or less independent of the morphological and syntactic structure of the sentence, is a useful and convenient instrument of expressing its theme-rheme structure, inherited from older linguistic stages. The language **keeps it because of its functional value**, in spite of the syntactic ambiguity it allows for. This disproves, at least partially, Jespersen's reducing free word order to "disorder"<sup>22</sup> and his claim that "the tendency towards a fixed word order" is "a progressive one"<sup>23</sup>.

At the end, we should emphasize the **diachronic and structural specificity of the syntagmatic type of Modern Bulgarian**. It has developed from a congruence language with a large all-Slavonic case system into a language with a reduced New Balkan case system<sup>24</sup>, and then into a congruence language which has lost its nominal cases but has preserved its relatively free word order.

## References

- Aikhenvald 2000: A i k h e n v a l d , A. Y. *Classifiers: A Typology of Noun Categorization Devices*. Oxford University Press, 2000.
- Andréasson 2001: A n d r é a s s o n , D. *Active languages*. Thesis for the degree of Bachelor of Arts in General Linguistic. Supervisor: Maria Koptjevskaja-Tamm. Department of Linguistics, Stockholm University, 2001. (pdf)
- Avgustinova 1994: A v g u s t i n o v a , T. On Bulgarian verbal clitics. – *Journal of Slavic Linguistics* 2(1), 29–47.
- Belić 1938: B e l i ć , A. O jezičkoj prirodi i jezičkom razvitku. – *Lingvistička ispitavanja*, knj. 1. Beograd, 1938, 103–104.

---

<sup>21</sup> About the communicative theme and the grammatical subject in English cf. Michael Halliday (1970).

<sup>22</sup> Jespersen (1969: 360).

<sup>23</sup> Jespersen (1969: 364).

<sup>24</sup> This is so in 17-th century Early Modern Bulgarian (Geshev 2007).



- Bickel 2005: Bickel, B. Typology in the 21st century: major current developments. *21century\_typology\_bickel\_submitted.pdf*. [visited September 2016] [https://en.wikipedia.org/wiki/Linguistic\\_typology](https://en.wikipedia.org/wiki/Linguistic_typology)
- Bloomfield 1970: Bloomfield, L. *Language*. London: George Allen & Unwin Ltd, 1970 (1933).
- Bogusławski 1977: Bogusławski, A. *Problems of the thematic-rhematic structure of sentences*. Warszawa: Państwowe wydawnictwo naukowe, 1977.
- Dyer 1992: Dyer, D. *Word Order in the Simple Bulgarian Sentence: A Study in Grammar, Semantics and Pragmatics*. 1992. Amsterdam: Rodopi Editions, B.V., 1992.
- Geshev 1990: Gešev, V. Analitizām, bezpadežnost, izolacija, degramatikalizacija. – *Bălgarski ezik*, XL, 1990/4, 293–306.
- Geshev 1992: Gešev, V. Njakoi osobenosti na predloga kato svoeobrazna čast na rečta. – *Săpostavitelno ezikoznanie/Contrastive linguistics*, 1992, 4, 32–44.
- Geshev 2007: Gešev, V. Rannonovobălgarskata padežna sistema. – *Palaeobulgarica / Starobălgaristika*, XXXI (2007), 1, 40–64.
- Geshev 2013: Gešev, V. Congruence languages and word order. – *Rasprave. Časopis Instituta za hrvatski jezik i jezikoslovlje*. Knjiga 39/2. Zagreb, 2013, 377–390.
- Halliday 1970: Halliday, M. A. K. Language structure and language functions. – In: *New Horizons in Linguistics*. Ed. by J. Lyons. Harmondsworth: Penguin, 1970. 140–165.
- Haspelmath 2001: Haspelmath, M. The European linguistic area: Standard Average European. – In: *Language Typology and Language Universals. Sprachtypologie und sprachliche universalien. La typologie des langues et les universaux linguistiques*. Vol. 2. (Ed. M. Haspelmath, E. König, W. Oesterreicher, W. Raible) Berlin–New York: Walter de Gruyter, 2001, 1492–1510.
- Ivančev 1978: Ivančev, Sv. *Prinosi v bălgarskoto i slavjanskoto ezikoznanie*. Sofia: Nauka i izkustvo, 1978.
- Jespersen 1969: Jespersen, O. *Language, its nature, development and origin*. London: George Allen and Unwin, 1969 (1922).
- Kibrik 2013: Kibrik, A. A. Peculiarities and origins of the Russian referential system. – In: *Languages across boundaries. Studies in memory of Anna Siewierska*. Eds. D. Bakker, M Haspelmath. Berlin / Boston: Walter de Gruyter, Mouton, 2013, 227–262.
- Klimov 1974: Klimov, G. A. On the character of languages of active typology. – *Linguistics*, 131, 1974, 11–25.
- Klimov 1977: Klimov, G. A. *Tipologija jazykov aktivnogo stroja*. Moskva: Nauka, 1977.
- Leafgren 2002: Leafgren, J. *Degrees of Explicitness: Information Structure and the Packaging of Bulgarian Subjects and Objects (= Pragmatics and Beyond New Series 102)*. Amsterdam: John Benjamins, 2002.
- Lehmann 1976: Lehmann, W. P. From Topic to Subject in Indo-European. – In: *Subject and Topic*. Ed. by C. N. Li. New York: Academic Press, 1976, 447–456.
- Lehmann 2002: Lehmann, Ch. *Thoughts on grammaticalization*. Second, revised edition. (Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt Nr. 9). Erfurt Seminar für Sprachwissenschaft der Universität, 2002; <http://www.uni-erfurt.de/sprachwissenschaft/ASSidUE/ASSidUE09.pdf>
- Lyons 1968: Lyons, J. *Introduction to theoretical linguistics*. Cambridge: Cambridge University Press, 1968.

- Papageorgiou 1985: Papageorgiou, G. *Novogrăcka gramatika*. Sofia: Nauka i izkustvo, 1985.
- Penčev 1980: Penčev, J. Aktualno delenie i intonacija. – In: *Sbornik praci Filozofické fakulty Brněnské univerzity. Studia minora Facultatis philosophicae Universitatis brunensis*, A 28, 1080, 195–204.
- Popova 2012: Popova, M. *Semantični i pragmatični aspekti na sintaktičnite kategorii v sāvremennija ispanski ezik*. Sofia: Universitetsko izdatelstvo Sveti Kliment Ohridski, 2012.
- Sgall et al. 1973: Sgall, P., E. Hajičová, E. Benešová. *Topic, focus and generative semantics*. Forschungen Linguistik und Kommunikationswissenschaft 1. Kronberg Taunus: Scriptor Verlag, 1973.
- Sgall et al. 1980: Sgall, P., E. Hajičová, E. Buráňová. *Aktuální členění věty v češtině*. (Československá akademie věd. Studie a práce lingvistické 12). Praha: Academia, 1980.
- Sintaksis 1983: *Gramatika na sāvremennija bălgarski knižoven ezik. Tom III. Sintaksis*. Ed.: K. Popov. Sofia: Izdatelstvo na Bălgarskata akademija na naukite, 1983.
- Skalička 2004: Skalička, V. Poznámky ke kongruenci (Bemerkungen zur Kongruenz, 1937, překlad A. Šimečková). *Souborné dílo*. Ed. Fr. Čermák, J. Čermák, P. Čermák, C. Poeta. Praha: Nakladatelství Karolinum, 2004, 168–171.
- Tisheva 2014: Tiševa, J. *Pragmatika i ustna reč*. (Kak govori sāvremennijat bălgarin, tom 2). Sofia: Fond naučni izsledvanija, Fondacija Fokus, 2014.
- Tisheva, Dzhonova 2006: Tiševa, J., M. Džonova. Starijat “nov” topik. – In: *Slavistika i obštestvo*. Ed. M. Mladenova, I. Monova. Sofia: Heron Press, 2006, 231–237.
- Uhlířova 2011: Uhlířova, L. Obštnost i mnogoobrazie na slovoreda na klitikite (Sāpostavitelno izsledvane na bălgarskija i češkija ezik). Čast I; II; III. *Sāpostavitelno ezikoznanie/Contrastive Linguistics*, XXXVI, № 1, 5–17; № 2, 20–31; № 3, 5–16.

### *За синтагматичната типология на съвременния български език*

Валентин Гешев

Софийски университет „Св. Климент Охридски“

В статията се въвежда понятието *синтагматична типология*, което отразява взаимодействието на различни механизми за синтагматично свързване на думите в изречението като словоред, съгласуване, формални падежи, лексикално значение, одушевеност, формални показатели на частите на речта, прозодия, определеност, удвояване на допълнението, извънезиково знание. Въз основа на това взаимодействие се абстрахира синтагматичният тип на *съгласувателните индоевропейски езици в Европа*, чиито представители са славянските езици, романските (без същинския френски), индоевропейските езици на Балканите, континенталният германски (немско-нидерландският езиков континуум). Съвременният български език е образцов представител на безпадежния подтип на съгла-

свателните индоевропейски езици в Европа, като в него съгласуването е важен инструмент за запазване на сравнително свободния комуникативно натоварен словоред след исторически засвидетелстваното драстично съкращаване на падежната система в него (общославянска в старобългарски и среднобългарски, общобалканска в ранноновобългарски и нулева при имената в съвременния наддиалектен разговорен български). Представени са примери, които показват, че синтагматични механизми като одушевеността, определеността, удвояването на допълнението не могат еднозначно да различат подлога на изречението от прякото допълнение в него. Единственият механизъм, който ги различава еднозначно, е съгласуването между подлога и сказуемото, но и той действа само ако безпредложните именни групи, изразяващи подлога и прякото допълнение, са в различно число (и евентуално от различен граматичен род). В противен случай на български са възможни синтактично двусмислени изречения и носителите на езика понякога съзнателно използват тая негова възможност.

Ако се опитаме да отговорим на въпроса защо в съвременния български (и в други езици с подобна синтагматична типология) няма тенденция темата в изречението да се преобразува в негов разположен преди глагола синтактичен подлог (както в английския език), можем да предположим че:

а) Напълно двусмислените изречения (за чието разгадаване не помага нито извънезиковото ни знание, нито конкретният контекст) са пренебрежимо малко статистически.

б) Носителят на езика може да избегне двусмислието, като смени комуникативната си стратегия и използва например синтактично еднозначни емфатични или страдателни конструкции.

в) Сравнително свободният словоред, повече или по-малко независим от морфологичната и синтактична структура на изречението, е полезен и удобен инструмент за изразяване на темо-ремната му структура, наследен от предходни езикови състояния. Езикът го пази поради високата му функционална стойност въпреки синтактичната двусмисленост, която той би могъл понякога да предизвика.

e-mail: vgeshev@slav.uni-sofia.bg  
St. Kliment Ohridski University  
15 Tzar Osvooboditel Blvd.  
Sofia 1504