

Towards the Definition of the Bulgarian Word Order System

Anton Zimmerling

Pushkin State Russian Language Institute; Institute of Linguistics Russian Academy of
Science, Russia

Abstract The world's languages with clitic clusters pattern with four major types of the word order systems, depending on the presence or absence of the 2P condition and the parameter of clitic-verb adjacency. Bulgarian has a double-focus system of clitic placement (2P condition and clitic-verb adjacency), which has typological parallels outside Europe but lacks direct counterparts in modern European languages. Neither the analogy with standard 2P languages without clitic-verb adjacency nor the analogy with the Romance systems with vP-internal clitics captures the profile of the Bulgarian clitic syntax. Historically, the rise of the clitic-verb adjacency is an innovation of Bulgarian, but its exact time and triggers are unclear. The language of the Wallachian letters (ca. 1386-1509) has a marked tendency towards the clitic-verb adjacency and is typologically similar to Modern Bulgarian but still has residual #XP – CL – [Y] – V orders. This idiom spoken by the L2 speakers of Middle Bulgarian cannot be viewed as a source of the Modern Bulgarian but hints that the clitic-verb adjacency parameter could develop in the history of Bulgarian because of the contact influence on the part of some Non-Slavic clitic systems.

Keywords Typology. Word order. Clitics. Clitic clusters. Bulgarian. Syntax. Prosody.

Summary 1 Ordering of Non-clitic Categories. – 1.1 Greenbergian Typology. – 1.2 Sentence Cartography. – 2 Clitics and Word Order Typology. – 2.1 Wackernagel's Law and Tobler-Mussafia's Law. – 2.2 Clitic Hosts and Clausal Architecture. – 2.3 Clitic Clusters and Clustering Clitics. – 3 Word Order Systems with Clustering Clitics. – 4 Defining the Bulgarian Word Order System. – 4.1 Synchrony. – 4.2 Diachrony and the Inventory of Clitics. – 4.3 Old Church Slavonic and Vernacular Texts. – 5 Middle Bulgarian: The Wallachian Letters. – 6 Conclusions.



Edizioni
Ca' Foscari

Peer review

Submitted	2021-10-13
Accepted	2021-12-22
Published	2022-02-15

Open access

© 2021 | Creative Commons Attribution 4.0 International Public License



Citation Zimmerling, A. (2021). "Towards the Definition of the Bulgarian Word Order System". *Balcania et Slavia*, 1(2), 239-256.

1 Ordering of Non-clitic Categories

Word order systems can be classified in terms of linearisation principles. For the ordering of non-clitic categories, two different measures – branching conditions and sentence cartography – are available.

1.1 Greenbergian Typology

The approach introduced by Greenberg (1963) is based on branching conditions, i.e. relative ordering of the head (X^0) vs complement in binary groups (A⁰, B) and the so-called sentence formula or basic word order showing the ordering of three diagnostic sentence categories – subject (S), object (O) and verb/predicate (V): in the pair (V⁰, O), the verb is analysed as the head; in the pair (S⁰, V), the subject under standard assumptions is analysed as the head. In the early versions of the Greenbergian typology (cf. Hawkins 1983), all languages including the so-called scrambling (free word order) languages, where the phrasal categories can be ordered in more than one way, have the basic word order per definition. With this approach, Slavic languages are relatively uniform: they can be classified as SVO languages with a predominant right branching. The modern data-oriented version of the Greenbergian analysis, based on the World Atlas of Language Structures (WALS) sample of 1376 world's languages (Dryer 2013b), confirms the existence of all six logically possible combinations of S, V, and O, i.e. SOV, SVO, VSO, VOS, OVS, OSV, although the last three types are rare (2,9% of the sample). The seventh class in Dryer's classification – languages lacking a dominant word order (189 languages, 13,73%) – is rather a concession to the agnostic view than a challenge to the initial Greenberg's assumption that all languages have a basic word order. I interpret it as a technical requirement that all languages in a typological sample must be described at the same depth of analysis: if a basic word order in a European language like German, Dutch, Frisian, Hungarian, Modern Greek, Welsh Romani, or Belorussian can only be established based on a formal syntactic analysis, which has not been done for other languages from the sample, the language patterns with the class lacking a basic word order.¹ A further problem with the input data is that typological databases occasionally put languages with similar word order systems into different classes because of the decisions made by the authors of the ref-

I am grateful to the anonymous reviewers for the valuable comments. The responsibility for all shortcomings is on the Author.

1 The notion of 'depth' is discussed in Haspelmath 2019.

erence grammars. For example, Modern Belorussian is classified in Mayo (1993, 924) as a language lacking a basic word order and allowing both VO and OV,² while Modern Russian is classified as an SVO, VO language (Bivon 1971) in a description made 22 years earlier. It is unclear whether such decisions are motivated by the real differences between the languages, by the theoretical sympathies of the authors, or by the changing beliefs in the linguistic mainstream.³

1.2 Sentence Cartography

Another approach for ordering sentence categories has been known as ‘template analysis’ or ‘sentence cartography’. The insight behind it is that there is a correlation between word classes/types of category and their linear position so that each category has its diagnostic slot in the template (‘map’) of a sentence. Sentence templates can be interpreted both as language-specific rules and as a manifestation of universal ordering principles. Early versions of the template analysis arose as empiric generalisations on the syntax of selected Germanic languages with a presumably fixed order of the post-finite categories (Diderichsen 1946), while classical versions of the 1990s-2000s combine the template analysis with the hypothesis on generalised phrase structure (Cinque 1999; 2014; Belletti 2004). The cartographic approach is compatible with the analysis of the word order variation. However, if a category alternates between two or more slots P_1, P_2, \dots, P_n , sentence templates become bulky since the alternations add extra slots, cf. the discussion in Zimmerling (2013, 28-33). Therefore, cartography is generally associated with those zones in clausal structure, where the ordering is more rigid, e.g. the clausal left periphery (Krapova 2002; Benincà, Munaro 2011), or clitic-internal ordering.

2 Clitics and Word Order Typology

The term ‘clitic’ emerged as a generalisation of the ancient terms *enclitic* ($X=CL$) and *proclitic* ($CL=X$), i.e. weak stress elements adjoining to their prosodic hosts (stressed words) from the right or the left, respectively (Spencer, Lúis 2010). It has been introduced to language

2 The WALS sample for the V^0, O languages includes 1516 languages, 101 of which (6,59%) are classified as “languages with both VO and OV” (Dryer 2013c).

3 As for Modern Bulgarian, all sources cited in Dryer 2013b and 2013c. Cf. Hubenova et. al. 1968, 31; Scatton 1984, 374; Dyer 1992, 16, 55, 149, who classify it as SVO, VO language.

theory by Arnold Zwicky (1977), who claimed that prosodically deficient elements incapable of making a phonetic word without combining with other elements (= ‘phonetic clitics’) are also syntactically deficient, i.e. have a special distribution in syntax and take positions not available for the non-clitic categories.⁴

2.1 Wackernagel’s Law and Tobler-Mussafia’s Law

The first observations that clitics constrain the clausal syntax were made in the late nineteenth century by Berthold Delbrück and Jacob Wackernagel. The latter showed that in some Old Indo-European languages including Old Greek, Old Indian, Avestan, Old Persian fixed-position enclitics took the clausal-second position after the first stressed word and proved that this feature of the second-position clitics (2P clitics) is triggered by a general principle, which does not depend on the syntactic type or the origin of the clitic element (Wackernagel 1892). The discovery of Hittite and Luwian confirmed that Wackernagel’s law was characteristic of most Old Indo-European languages. A definitional property of 2P clitics is that they attach to at least two different syntactic types of the host (X, Y), but do not license them simultaneously:

(i) #X=C, #Y=CL, *#X, Y-CL

The so-called Tobler-Mussafia’s law (Tobler 1875; Mussafia 1886) is based on the distribution of Old French and Old Italian clitic pronouns: they are verb-adjacent and license the orders V=CL ~ CL=V clause-internally, but not clause-initially, *#CL-V. This is a consequence of Wackernagel’s law, preventing strict enclitics from taking the clause-initial position. Similar conditions are attested in Old Spanish (Wanner 1996). Bulgarian predicate clitics (clustering pronouns and auxiliaries) have the same bundle of features: they are both verb-adjacent and clause-internal. For such languages, it is possible to treat Tobler-Mussafia’s law as a non-syntactic condition imposed on the parameter of the verb-clitic adjacency (cf. Franks 2008). This conclusion is prompted by the fact that Old Romance word order systems are historically transitional from systems with Wackernagel’s law and 2P clitics to systems of the Modern Romance type with verb-adjacent clitics. Tobler-Mussafia’s law is redundant for the

⁴ The initial version of Zwicky’s analysis had a stipulation that special syntactic distribution is characteristic only of a subclass of clitics labelled “special clitics”, but not for the residue labelled “simple clitics”. However, simple clitics are elusive since their recognition crucially depends on the chosen theory of sentence derivation.

description of those languages with 2P clitics, which lack the parameter of obligatory clitic-verb adjacency and license configurations like #X=CL ...V.

2.2 Clitic Hosts and Clausal Architecture

Roman Jakobson (1935) was the first linguist who admitted that Wackernagel's law is instrumental in Modern Slavic languages, notably in the Balkan Slavic group, which partly retains the old accent conditions. Andrej Zaliznjak (1993) was the first linguist who proved that vernacular Old Russian had Wackernagel's law and explained, how the old East Slavic clitics were eliminated in the subsequent history of Russian (2008). Both Jakobson and Zaliznjak associated Wackernagel's law with 2P clitics after the first phonetic word and claimed that Modern Slavic languages placing 2P clitics after the first constituent (Slovenian, Bulgarian, Czech, Slovak, etc.) lost Wackernagel's law. This claim finds little support from linguistic typology since Wackernagel's law is not a unique feature of Old European languages. From the 1920s on, 2P clitics were discovered in many genetically not related world's languages from different areas, including Luiseño and Mayo (Uto-Aztecan), Ossetic, and Pashto (Indo-Iranian), Warlpiri and Djaru (Pama-Nyungan), etc. A part of these languages, e.g. Ossetic, Pashto, Kavineña (Tacanan), Kashibo-Kakataibo (Pano) places clausal 2P clitics after the first spelled-out constituent, not the first phonetic word, while some languages, e.g. Bosnian-Croatian-Serbian (BCS), Luiseño and Warlpiri license the first phonetic word/first constituent variation (#X/XP=CL) (cf. Halpern 1996). Finally, clausal 2P clitics are also attested in languages that lack initial NPs and initial multi-word constituents, e.g. in Lummi (Salish), or Kabyle (Afroasiatic).

In subordinate clauses, 2P clitics generally attach to the complementizer (Comp=CL), though languages where the clitics skip the complementizer and attach to the next category Y (Comp Y=CL) are attested as well. The latter scenario is required in Pashto (2P, no verb-clitic adjacency), Tagalog (2P & verb-clitic adjacency) and optional in colloquial Czech (2P, no verb-clitic adjacency) and Bulgarian (2P & verb-clitic adjacency) (cf. Zimmerling 2013, 64, 114, 437). Additional types of the clitic hosts licensed by some languages with 2P clitics include initial proclitics and sentential complements ($[_{\text{CP}}\dots]=\text{CL}$): the latter option is regularly used in Cavineña, Kashibo-Kakataibo, and Slovenian, sporadically also in Czech, Upper Sorbian, South-Eastern Tepehuan, BCS, and Bulgarian (Zimmerling 2013, 154-62, 434-8). The principle of 2P clitic placement is strikingly uniform across the world's languages and compatible with different morphosyntax, while the mechanisms licensing diverse clitic hosts in

languages with 2P depend on the clausal architecture in the corresponding language. This makes the 2P syntax an interface phenomenon: a linguist must both explain the locus of the clitics, and the choice and internal structure of the hosts. The terms ‘prosodic’ and ‘syntactic’ used in two competing descriptions of the same data – the situation in BCS, a language with 2P clitics and the #X=CL ~ XP=CL variation (Radanović-Kocić 1996; Progovac 1996) – are misleading: they rather refer to different models of the prosody-syntax interface than to the possibility to eliminate all syntactic vs prosodic information from the description.

2.3 Clitic Clusters and Clustering Clitics

A cluster is a complex object satisfying the following criterion:

- (ii) If a syntactic domain contains two (or more) clitics *a* and *b* and they stay contiguously, they are placed in a rigid order $a > b$ (‘*a* immediately precedes *b*’).

The clitics conforming to the criterion (i) are called *clustering*, while the rules predicting clitic-internal ordering are called *clitic templates* or *ranking rules* (Zaliznjak 1993, 282; Franks, King 2000). True clusters must be distinguished from occasional sequences of phonetically adjacent clitics. The latter can belong to different hierarchically independent syntactic domains, while clustering clitics represent one domain.⁵ On the synchronic level, clitic templates are idiosyncratic since the template order cannot be derived from the order of non-clitic categories. However, it is up to a certain extent possible to explain it historically (Zimmerling 2012) and to reconstruct the principles of the prosody-to-syntax interface underlying the templates (Billings, Konopasky 2002). For Slavic languages, it has been shown that despite the clusters often contain clitics of more than one type – pronouns vs auxiliaries vs particles – no Slavic language licenses the insertion of an auxiliary into a block of clitic pronouns or vice versa. Other factors such as cliticization time or prosodic weight do not override in the Slavic languages the principle of grouping the clustering elements into blocks of clitics sharing the same syntactic category (Franks 2008; Zimmerling, Kosta 2013).

Clustering clitics are attested both in 2P languages and in languages with vP/VP-internal clitics. At the same time, there exist lan-

⁵ In languages with the parameter of clitic climbing, e.g. in Slovak and BCS, the clitic template of the matrix clause includes slots for the climbed clitics generated in the embedded clause (Zimmerling 2013, 137).

languages with single non-clustering 2P clitics. The ability to form a cluster is not an inherent characteristic of clitics: a clitic morpheme can be clustering in some configurations and non-clustering in other configurations in the same language. In Slavic languages, only clause-level clitics cluster. For example, Old Church Slavonic, Old Russian, and BCS accusative clitics are clustering as clause-level argument clitics but non-clustering as PP-level clitics, when they attach to prepositions. Old Russian free clitic же ‘emphatic marker’ was clustering as an emphatic sentential particle, but non-clustering as a non-sentential additive marker ($X \text{ же } Y$ ‘X and Y’), or as an identifier (Zaliznjak 1993, 281) (for the external parallels from Non-European languages, cf. Zimmerling 2012). The same holds for Bulgarian dative clitics: they are clustering as clause-level argument clitics but non-clustering as possessive DP-level clitics.⁶ Such facts support the claim that clusterization takes place in dedicated syntactic positions and is blocked in other positions. A possible solution to this puzzle is to analyse the clusters as phrases of a special kind, CLP (Zimmerling, Kosta 2013). A single clustering clitic is interpreted as an occurrence of CLP, while the number of clitics in a sentence depends on such factors as the valency frame of the predicate, the type of the clause, etc.

Some authors argued that all parts of the cluster must be prosodically homogeneous (Franks 2008, 95), but there is empirical counter-evidence against this claim. Clustering and non-clustering clitics have non-identical properties. Clusters arise as rhythmic patterns and their elements often represent different layers of cliticization, therefore some clustering elements can lack full properties of phonetic clitics. This was arguably the case with 1-2 p. present tense indicative BE-auxiliaries in Old Russian (Zaliznjak 2008, 37). Nevertheless, the ordering of these elements was rigid (225-7).

3 Word Order Systems with Clustering Clitics

Languages with clustering clause-level clitics pattern with four types depending on the presence/absence of the clitic-verb adjacency and the 2P principle constraining either the position of the clustering clitics (CL) or the position of the verb (V). Henceforth, I apply the notation introduced in Zimmerling 2013 and Zimmerling, Kosta 2013. The position of the pivotal category (CL or V) respective the clausal left border (#) is defined in terms of surface syntax, while the preceding 2P is analysed as X/XP, i.e. intra-clausal position licensing at least two

⁶ It is dubious that the suffixed article and the possessive pronoun in Bulgarian DPs like [_{DP} [книга-та]=му] represent the same level of cliticization.

different categories A v B, but not simultaneously (Zimmerling 2015).

iii $[_{XP} A v B] - 2P, *[_{XP} A \& B] - 2P$

The tag 'W-system' ('W' reads either 'word' or 'Wackernagel') stands for word systems with 2P clitics lacking the parameter of clitic-verb adjacency. W-systems are subdivided into W_1 -systems, i.e. languages lacking clause-initial NPs and W_2 -systems licensing them. W_2 -systems are further subdivided into W_{2A} -systems, where 2P clitics are placed after the first phonetic word and split the initial multi-word constituent ($\#[_{XP} W^1 = CL, W^2..W^n]$), W_{2B} -systems, where 2P clitics are placed after the first spelled-out constituent and W_{2C} -systems, which license the $X/XP = CL$ variation. W_2 -systems are attested in different areas and are common, while W_1 -systems are only attested in those verb-initial languages, where V1 alternates with an operator category (clause-initial quantifier or T(ense) A(spect) M(ood) marker).

The tag 'W⁺-system' ('W⁺' reads 'modified W')⁷ stands for word order systems with clitic-verb adjacency and the 2P condition defined for one of the pivotal categories – either CL or V. There are two main varieties. In the Philippine-type W⁺-systems, the pivotal category is CL, which takes clausal 2P. If clitics are present, the verb adjoins to them either from the left or from the right ($\#V-CL \sim \#XP-CL-V$). Word order systems of this subtype are characteristic of Austronesian languages, especially from the Central Philippine area. Bulgarian seems to be the only representative of this subtype in Modern Europe. In the Germanic-type W⁺-systems, the pivotal category is V, which takes the V2 or the V1/V2 positions, while the clitics are realised in the postverbal positions as CL2/CL3 depending on the position of the verb. Word order systems of this type are attested only in Old Germanic languages: the V2 and V1/V2 constraints are typologically rare outside Europe.

The tag W*-system ('W*' reads 'degraded W') stands for languages, where the 2P placement remains an option for the clause-level clitics and/or is not generalised for all types of clauses. This is, for example, attested in Old Church Slavonic, Modern Polish, Gurindi, and Mudburra (both are Pama-Nyungan languages).

Finally, the tag V-system ('V' reads 'verb') stands for languages of the Modern Romance type, where neither V nor CL has a fixed position respective to the clausal left border in the diagnostic type of clauses, but the verb and the clitics are adjacent ($\#...[V + CL]...$). V-systems constitute a majority of European word order systems with the clustering clitics, but exact parallels outside Europe are rare.

⁷ The term 'W⁺-system' was first coined by the author of this article in 2002.

Table 1 Word order systems with clustering clitics, after Zimmerling 2013

	W-systems				W ⁺ -systems		W [*] -systems	V-systems
	W ₁	W _{2A}	W _{2B}	W _{2C}	Philippine type	Germanic type		
clitic-verb adjacency	no	no	no	no	yes	yes	no	yes
2P condition for CL	yes	yes	yes	yes	yes	no	(yes)	no
2P condition for V	no	no	no	no	no	yes	(no)	no
Languages	Lummi, Kabyle	Hittite, Old Greek, Old Novgorod, Russian	Czech, Slovak, Slovenian, Ossetic, Pashto, Cavineña	BCS, Luiseño, Warlpiri	Tagalog, Cebuano, Bikol, Masbatenyo, Bulgarian, Old Italian	Old Icelandic, Middle Norwegian	Old Church Slavonic, Polish, Gurindi, Mudburra	French, Italian, Spanish, Rumanian, Albanian, Modern Greek

W-systems, W⁺-systems, and V-systems represent stable ordering principles, while W^{*}-systems are transitional, which partly explains the uneven coverage of these word order systems.

4 Defining the Bulgarian Word Order System

4.1 Synchrony

There are two main approaches to defining the Bulgarian word order system - to analyse it as a language with 2P clitics with an added CL-V adjacency (Gălăbov 1950; Dimitrova-Vulchanova 1999) or as a language with verb-adjacent clitics with a non-syntactic 2P condition (Tobler-Mussafia's law) imposed on the distribution of CL (Rå Hauge 1976; Franks 2008). The first solution is more popular in functional studies, the second one in formal descriptions of Bulgarian. Both approaches are complementary, but from the typological perspective, the first one is preferable since the features '+ 2P' and '+ verb-adjacent' do not exclude each other. The parallel between Bulgarian and Central Philippine languages was first acknowledged in Billings, Konopasky 2002, cf. also a survey of Philippine W⁺-systems in Lee, Billings 2005. The dichotomy of mutually exclusive taxons '2P clitics' vs 'verb-adjacent clitics' is prompted by the current distribution of W-systems and V-systems in modern Europe, whereas Bulgarian is the sole representative of the Philippine-type W⁺-system in this area. V-systems of the Modern Romance type, where V and CL are

adjacent, but neither V nor CL has a fixed position respective to the clausal left border, are rare outside Europe.

The orientation of Bulgarian clitics towards the left periphery is proved by such tests as the impossibility of the *NP1-NP2-V-CL order, cf. the pair of sentences, where both nominal arguments are indexed by the pronominal clitics, but the *#X_i-Y_j V-[_{CL} a_i b_j] order is blocked:

- (1) Blg. a. [На Иван]_i [книгата]_j =съм=му_i=я_j върнал
'I have returned the book to Ivan', lit. 'To Ivan, the book, I gave it back to him'.
b. *[На Иван] [книгата] върнал =съм=му=я.

Another test confirming the 2P-orientation of Bulgarian clustering clitics is based on the distribution of periphrastic verbal forms containing an optative marker (not a clustering clitic itself). These forms host the clitics, but do not combine with other elements in XP:

- (2) Blg. a. [купил бих] (1)=я_i книгата,
'I would buy that book'.
b. [книгата]_i (1)=я_i бих купил
'I would buy that book'.
c. *[книгата]_i (1) [купил бих] (2)=я_i.

4.2 Diachrony and the Inventory of Clitics

The exact definition of the Proto-Slavic clitic system is a matter of debate. Jakobson (1935) and Gălăbov (1950) assumed that Proto-Slavic had Wackernagel's law, i.e. it was a standard W-system in terms of this paper, more exactly a W_{2A}-system. This assumption is based on the fact that the core of the 2P clitic inventory - the particles же 'emphatic marker', ли 'yes-no marker' and the 1-2 p. dative pronouns ми '1Sg.Dat', ти '2Sg.Dat', си 'Refl.Dat', ны '1Pl.Dat|Acc', вы '2Pl.Dat|Acc', на '1Du.Dat|Acc', ва '2Du.Dat|Acc' - is common Slavic, and all Old Slavic languages use them as clustering 2P elements. The same holds for the 2P particle бо 'causal marker' preserved in South Slavic and Southern Old Russian (Zaliznjak 2008, 31). The short accusative forms ма '1Sg.Acc', та '1Sg.Acc', са 'Refl.Acc', и '3Sg.M.Acc', ю '3Sg.F.Acc', ю '3Sg.M.Acc', ъ '3Sg.N.Acc', ја '3Du.M.Nom|Acc' have common Slavic morphology and cluster clause-internally. However, they retain some non-clitic side uses in Old Church Slavonic texts, which shows that they were cliticised later than the 1-2 p. dative pronouns (Vaillant 1948, § 262; Zaliznjak 2008, 36).

In more recent times, the idea that Bulgarian inherited the 2P condition from Proto-Slavic was challenged. Pancheva (2005) argues that

not only the rise of the CL-V adjacency but also the uniform treatment of clitic pronouns and BE-auxiliaries as 2P clitics was an innovation realised during the history of Bulgarian. It is well-known that Slavic present tense indicative BE-auxiliaries represent a relatively recent layer of cliticization and were made part of the cluster later than dative and accusative pronouns. This is mentioned by Zaliznjak (1993, 285) regarding the positions of Old Russian 1-2 p. auxiliaries, which show the enclitic behaviour only as part of the perfect construction with the *-l* participle. The slots of auxiliary clitics in a cluster are different in three areal types of the Slavic template. Old East Slavic dialects place the 1-2 p. auxiliary clitics⁸ after the block of clitic pronouns in the slot labelled 'AUX2' in Zimmerling 2013, 110, 337.⁹

(iv) East Slavic [_{CLP} [_{PTCL}...]] [_{PRON}...] AUX2]

West Slavic languages place the auxiliary clitics before the block of clitic pronouns in the slot labeled AUX1 (Zimmerling 2013, 110, 337).

(v) West Slavic [_{CLP} [_{PTCL}...] AUX1 [_{PRON}...]]

Finally, Balkan Slavic languages including Bulgarian use both AUX1 and AUX2: most auxiliaries take AUX1, while the 3Sg =*je* takes AUX2 (Franks, King 2000; Zimmerling 2013, 111, 337).

(vi) Balkan Slavic [_{CLP} [_{PTCL}...] AUX1 [_{PRON}...] AUX2]

It is plausible that the AUX1 slot is an innovation compared to AUX2: vernacular Old East Slavic texts with clitic clusters and the active AUX2 slot date back to the eleventh century, while the earliest comparable West Slavic texts come from the first half of the fourteenth century. The migration of auxiliary clitics from AUX2 to AUX1 is attested in the fourteenth century in Old Serbian (Pavlović 2011) and Old Czech (Zimmerling 2013, 474-8). However, neither the observed geographical variation of the AUX slots nor the inconsistent placement of the auxiliaries in some Old Slavic dialects proves that Proto-Slavic lacked Wackernagel's law. Wackernagel's law predicts the 2P placement of all elements that belong to the class of clustering clitic-

⁸ Most Old Russian dialects lack overt 3 p. present tense auxiliary clitics. This holds for the North-Western (Novgorod and Pskov) and likely for the Western (Smolensk and Polotsk) dialects.

⁹ A historical continuation of the East Slavic template rule is represented in Carpathian Rusyn dialects (Tolstaya 2012). These dialects have extra slots for the new plus perfect auxiliaries.

ics. If this class is expanded during a period of time, it means that the same ordering principle is applied to a larger inventory of clitics. New clitics add extra slots to the existing templates or fill in the existing slots, but neither of these scenarios brings about a change of the word order system.

4.3 Old Church Slavonic and Vernacular Texts

Clitic clusters represent rhythmic patterns. One can disregard the type of the text only in a situation, when such patterns and word orders with clusters are generalised in all forms of a language, e.g. in the present-day codified Slavic languages. The situation in Old Slavic languages is less evident. Wackernagel's law and the conditions of standard W_2 -systems are best displayed in the texts close to the oral speech, like birch bark letters and the direct speech fragments in the Old Russian languages. This has been proved in detail by Zaliznjak (2008), who demonstrated that the inconsistent bookish Old Russian W^* -system was derived from the consistent colloquial Old Russian W -system by a parametric change licensing the end position of the clitics and the choice of the hosts. This change of parameters has both a diachronic and a stylistic dimension: the authors could shift the register in some hybrid text genres, e.g. in Old Russian chronicles. Letters and diplomas written by the professional scribes are generally a reliable source, too, especially in a situation, where non-bookish vs bookish texts require different forms of a language if not different languages (Church Slavonic vs vernacular Slavic idiom). Meanwhile, Old Church Slavonic as liturgic language poses a problem even for the group of Old South Slavic dialects most close to it: in those genres, where Old Church Slavonic was the required standard, the authors modelled their word order after the precedent texts, not their everyday speech. The same presumably holds for the early Old Church Slavonic, since the translators of the Slavic Bible in many cases copied the word order and clitic positions from the Greek text (Vaillant 1948, 261).

5 Middle Bulgarian: The Wallachian Letters

The word order system of the Wallachian letters (ca. 1386-1509) is similar to the Modern Bulgarian W^+ -system, though it cannot be viewed as its direct source: these letters are written by the L2 speakers of Middle Bulgarian, who also spoke a form of Rumanian and probably other languages. The clustering clitics take 2P, the contact orders of CL and V prevail. Still, there are 51 examples of the order #XP-CL-[Y]-V, with non-clitic categories intervening between

CL and V.

- (3) Middle ω сем же віе добръ знаете, како =ви =смь до нинъ¹ азъ² защитил
Blg. ω т Турци, [LVI: Vlad I Dracul. (c. 1431-1446)].¹⁰
'And you know it for sure, how I² have defended you from the Turks till now¹.

Examples like (3) allow to qualify the language of Wallachian letters as a W-system and not as a W⁺-system like Modern Bulgarian or Tagalog. Another deviation from Modern Bulgarian is that some authors licensed the clustering clitics clause-initially. This is attested a dozen times in letters nos. 74, 179, 212, 213, 215, 216, 226: in 3 examples the CL1 order comes up in yes-no questions like (4), in 8 examples it is attested after the initial proclitic и 'and', cf. (5) and in one example, cf. (6) the clitic is likely used after a prosodic break.

- (4) Middle Тои не знам: е=ли=ви доишла книга въ рѣце, или не е. [CCXII: Petru Uroş, trimesul Braşovenilor la împăratul Sigismund, şi Ştefan, logofătul lui Vlad Dracul. (c. 1432-1437)].¹¹
'I am not aware of that: did the book reach your hands, or not.'
- (5) Middle Казахъ=ми шни: шста Мартин, да доде съ дроуги товар. И=ми=
Blg. поръчи съ Ханеш. ако ест хтѣніе ти, пусти ми съ Ханеш дроуги дукати, понеже ест зде и ощъ мѣд, да я купѣ. [CCXIII: Gherghe Lascar, cămăraşul lui Vlad I Dracul. (c. 1437)].¹²

10 An anonymous reviewer indicates that word orders like (3) are marginally possible with some adverbials in some styles of Modern Bulgarian.

11 The Wallachian Bulgarian example (4) likely has contrastive stress on the initial auxiliary *e*. In this case, the syntax of (4) is compatible with Modern Bulgarian sentences like Blg. *Сте= ли или не=сте начело на движение за сединение на Княжество-то с Източна Румелия?* I am grateful for the anonymous reviewer for providing this example and for the discussion. The fronting of the auxiliary clitic in contexts, where it gets contrastive stress can be explained in a twofold way: a) prosodically - the deficient elements that are used as enclitics clause-internally are boosted and 'repaired' in contrastive contexts, b) information-structurally - the clustering clitics that normally take 2P, are fronted if the 1P order has communicative motivation (Zimmerling 2015). It should be mentioned that standard definitions of clitics generally specify that true clitics cannot take contrastive stress, be negated or conjoined. While this condition holds for the majority of word order systems with clustering clitics, some languages have an option for the early placement of clitics in certain non-canonic positions.

12 An anonymous reviewer points out that similar examples, where the clustering clitics occur in such initial or quasi-initial positions can be found in Modern Bulgarian texts as well, cf. Blg. *И той се=нахвърля върху ми. А ми= е жал за момчето!* Given that the variant [?]*А жал= ми=е за момчето!* is rated as odd, these facts suggest that Modern Bulgarian and the language of Wallachian Bulgarian letters have an option for the emphatic fronting of (some) clustering clitics. Similar type of clitic fronting is attested in Slovenian, which is a W_{2B} - system without the clitic-verb adjacency. The default order

‘They told me that Martin stayed there and is going to come with more goods. **And** send *me* a message with Hanes: “If you like, send me more ducats with Hanes, because there is more honey here so that I buy it”’.

- (6) Middle Друго, **ми**=дръжѣт Крстѣ Рошіул ѡнеи книге, не щет да=**ми**=**их**
Blg. даст; [CXXVI: Dragomir Udriște, vornicul lui Vlad III Călugărul. (c. 1482-1492)].
‘Second, Krstia Roshuil keeps these books **from me** and is not going to give *them to me*’.

The authors of the examples (4)-(6) likely lacked the Tobler-Musaffia’s law in their idioms, but there are no sufficient grounds to project this feature to Middle Bulgarian in general. The dialectal variation in the corpus of the Wallachian letters is not unlikely. Uroș (1432-37), Gherghе Lascar (1437), and Dragomir Udriște (1482-1492) or their scribes could have CL1 in their Middle Bulgarian dialects and/or in their other language (e.g. Old Rumanian). However, the contexts of examples (4) and (5) are not entirely telling for the comparison with Modern Bulgarian since they involve such pragmatic mechanisms as contrastive stress on the perfect auxiliary *e* in (4) and emphasis in (5): in these contexts, Modern Bulgarian occasionally or regularly licenses contrastive and emphatic fronting of the clustering clitics.¹³

Regarding the auxiliary slots, the language of the Wallachian letters shows an intermediate-type clitic template like Old Czech or Old Serbian, with two slots – AUX2 and AUX1 – available for most auxiliary clitics. The complementary distribution of the 3Sg =*e* (AUX2) vs all other present tense indicative BE-clitics (AUX1) diagnostic for the Modern Balkan type of the template has not yet established. Moreover, AUX2 is the main slot, while AUX1 is a reserved option. Past tense auxiliaries opened an extra slot AUX4 to the right from AUX2, cf. example (7). Similar templates are attested in some Modern Carpathian Rusyn dialects (Tolstaya 2012).

- (7) Middle И по сих давам у знаніе господству=**ви** израді ѡвою чловѣкъ мою
Blg. на име Стънислав, како=**му**=**ест**=**бил** дал¹ Марко един синъ=**му**, та да=**га** научит² ѡзык, да=**му** плащат; а ѡн, къда=**ест** бил³ летоска, ѡн=**ест** бежал⁴, тако=**ест** ѡтишел⁵ през Дунав и Хръсова та=**се**=**ест** потурчил⁶. [CCXLVII: Carstian, vornicul lui Radu IV cel Mare. (c. 1496-1507)].

for Slovenian clustering clitics is 2P, cf. Slv. *Videl=sem=ga* ‘I have seen it’. However, colloquial Slovenian has an additional 1P option, when the clustering clitics are fronted in verificational and emphatic contexts, cf. Slv. *#Sem=ga=videl* ‘I have indeed seen it’.

13 See the details in footnotes 11 and 12. I am grateful to the anonymous reviewer for the valuable comments.

'And now I notify your grace about my man, by the name Stānislav. Marko had previously sent¹ him one of his sons, so that Stānislav would teach² him the language for payment. But he ran⁴ away from where he was³ last summer, so he drove down⁵ the Danube from Hāršovo/Hârşova and converted⁶ to Islam [lit. became⁶ a Turk]'.
(7')

И	по сих	давам	у	знание	господству	=ви	
And	till now	gave:1SG	by	knowledge	dominion	you: 2PL	
израдї	ωвюу	чловѣкъ	мою	на	име	Стънислав,	
sake	that	man	mine	by	name	Stānislav	
како	=му	=ест	=бил	дал ¹	Марко		
how	him:CL.DAT	is: CL.3SG	was: 3SG	gave: 3SG	Marco		
един син	=му	та	да	=га	научит ²	язик,	
one son	him: CL	so	to	him: CL.ACC	teach	language	
да	=му	пощат;	а	ωн,	къда	=ест	бил ³
to	him: DAT.CL	pay: 3PL	but	he	when	is: CL.3SG	was: 3SG
летоска,	ωн,	=ест	бежал, ⁴	тако	=ест	ωтишел ⁵	
last.summer	he	is: CL.3SG	ran: 3SG	so	is: CL.3SG	departed: 3SG	
през	Дунав и	Хръсова	та	=се	=ест	потурчил ⁶	
Along	Danube and	Hāršovo	so	REFL.CL is: CL.3SG		became.a.Turk: 3SG	

[CCXLVII: Carstian, vornicul lui Radu IV cel Mare. (c. 1496-1507)].

'And now I notify your grace about my man, by the name Stānislav. Marko had previously sent¹ him one of his sons, so that Stānislav would teach² him the language for payment. But he ran⁴ away from where he was³ last summer, so he drove down⁵ the Danube from Hāršovo/Hârşova and converted⁶ to Islam [lit. became⁶ a Turk]'.
(vii)

The Middle Bulgarian template is shown schematically in (vi).

(vii) Middle Bulgarian [_{CLP} [_{PTCL}...]] AUX1 [_{PRON} DAT ACC] AUX2 AUX4]

6 Conclusions

Bulgarian has a double-focus system of clitic placement (2P condition and clitic-verb adjacency), which is not typologically rare but lacks exact counterparts in modern European languages. Therefore, neither the analogy with standard 2P languages without clitic-verb adjacency nor the analogy with the Romance systems with vP-internal clitics captures the profile of the Bulgarian clitic syntax. Bulgarian inherited the principle of 2P clitic placement and part of the clustering clitics inventory from Proto-Slavic, while the clitic-verb adjacency is an innovation. Its exact time and triggers are obscure, but a contact influence of a non-Slavic Balkan word order system is probable. The language of the Wallachian letters has residual #XP-CL-[Y]-V orders but is otherwise typologically similar to Modern Bulgarian. However-

er, this idiom spoken by the L2 speakers of Middle Bulgarian cannot be viewed as a direct ancestor of Modern Bulgarian.

Bibliography

- Belletti, A. (ed.) (2004). *Structures and Beyond. The Cartography of Syntactic Structures*, vol. 3. New York: Oxford University Press.
- Benincà, P.; Munaro, N. (2011). *Mapping the Left Periphery: The Cartography of Syntactic Structures*. New York: Oxford University Press.
- Billings, L.; Konopasky, A. (2002). "The role of morphology in ordering verb-adjacent clitics: from syntax to prosody in Bulgarian and Tagalog". Alexiadou, A. et al. (eds), *Linguistics in Potsdam*, vol. 19. Potsdam: Institut für Linguistik, Universität Potsdam, 1-26.
- Bivon, R. (1971). *Element Order*. Cambridge: Cambridge University Press. Studies in the Modern Russian Language 7.
- Bošković, Ž. (2004). "Clitic placement in South Slavic". *Journal of Slavic Linguistics*, 12(1-2), 39-90.
- Cinque, G. (1999). *Adverbs and Functional Heads. A Cross-Linguistic Perspective*. New York: Oxford University Press.
- Cinque, G. (2014). "Again on Tense, Aspect, Mood Morpheme Order and the 'Mirror Principle', Functional Structure from Top to Toe". *The Cartography of Syntactic Structures*, vol. 9. Oxford: Oxford University Press, 232-65.
- Diderichsen, P. (1946). *Elementær Dansk Grammatik*. København: Gyldendal.
- Dimitrova-Vulchanova, M. (1999). "Clitics in the Slavic languages". van Riemsdijk, H. (ed.), *Clitics in the languages of Europe* (Eurotype 20-5). Berlin; New York: Mouton, 83-121.
- Dryer, M.S.; Haspelmath, M. (eds) (2013a). *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dryer, M.S. (2013b). "Order of Subject, Object and Verb". Dryer, Haspelmath 2013. <http://wals.info/chapter/81>.
- Dryer, M.S. (2013c). "Order of Object and Verb". Dryer, Haspelmath 2013. <http://wals.info/chapter/83>.
- Dyer, D. (1992). *Word Order in the Simple Bulgarian Sentence*. Amsterdam: Rodopi.
- Franks, S. (2008). "Clitic Placement, Prosody and the Bulgarian Verbal Complex". *Journal of Slavic linguistics*, 16(1), 91-137.
- Franks, S.; King, T.H. (2000). *A Handbook of Slavic Clitics*. New York: Oxford University Press.
- Gălăbov, I. (1950). "Zur Frage der bulgarischen Enklitika". *Zeitschrift für slavische Philologie*, 10(2), 417-18.
- Greenberg, J.H. (1963). "Some Universals of Grammar with Particular Reference to the Order of the Meaningful Elements". *Universals of Language*. Cambridge (MA): MIT Press, 73-113.
- Halpern, A.L. (1996). "Introduction". Halpern, Zwicky 1996, ix-xxiii.
- Halpern, A.L.; Zwicky, A.M. (eds) (1996). *Approaching Second: Second Position Clitics and Related Phenomena*. Stanford: CSLI Publications.
- Haspelmath, M. (2019). "Ergativity and the Depth of Analysis". *Rhema. Рема*, 4, 118-30.
- Hawkins, J. (1983). *Word Order Universals*. New York: Academic Press.

- Hubenova, M.; Dzhumadanova, A.; Marinova, M. (1968). *A Course in Modern Bulgarian*, vol. 1. Columbus: Slavica.
- Jakobson, R. (1971). "Les enclitiques slaves". *Selected Writings*. Vol. 2, *Word and Language*. The Hague; Paris, 16-22.
- Krapova, I. (2002). "On the Left Periphery of the Bulgarian Sentence". *Working Papers*, vol. 12. Venezia: Università Ca' Foscari Venezia, 107-28.
- Lee, C.; Billings, L. (2005). "Wackernagel and Verb-Adjacent Clitics in Central Philippines". Heinz, J.; Ntelitheos, D. (eds), *Proceedings of the 12th meeting of the Austronesian Formal Linguistics Association (UCLA Working Papers in Linguistics 12)*. Los Angeles: Department of Linguistics, University of California, 241-54.
- Mayo, P. (1993). "Belorussian". Comrie, B.; Corbett, G.G. (eds), *The Slavonic Languages*. London: Routledge, 887-946.
- Mussafia, A. (1886). "Una particolarità sintattica della lingua italiana dei primi secoli". Avolio, C.; Biadene, L.; Ascoli, G.I. (a cura di), *In memoria Napoleone Caix e Ugo Angelo Canello: Miscellanea di filologia e linguistica*, 155-261. Firenze: Successori Le Monnier.
- Pancheva, R. (2005). "The Rise and Fall of Second-Position Clitics". *Natural Language and Linguistic Theory*, 23(1), 103-67.
- Pavlović, S. (2011). "Distribucija enklitika u starosrpskim poveljama i pismima XII i XIII veka" (The Placement of Enclitics in Old Serbian Documents of the XIIth and XIII Century). *Zbornik Matice srpske za filologiju i lingvistiku* (Matica Srpska Journal of Philology and Linguistics), LIV(2), 31-52.
- Progovac, L. (1996). "Clitics in Serbian/Croatian: Comp as the Second Position". Halpern, Zwicky 1996, 411-28.
- Radanović-Kocić, V. (1996). "The Placement of Serbo-Croatian Clitics: A Prosodic Approach". Halpern, Zwicky 1996, 429-48.
- Rå Hauge, K. (1976). "The Word Order of Predicate Clitics in Bulgarian". *Journal of Slavic Linguistics*, 7(1), 89-137. (Originally appeared as Meddelelser 10, Oslo: Slavisk-baltisk institutt, Universitetet i Oslo).
- Sadock, J.M. (1995). "A Multi-Hierarchy View of Clitics". *CLS 31*. Part 2, 258-79.
- Scatton, E.A. (1984). *A Reference Grammar of Modern Bulgarian*. Columbus: Slavica Publishers.
- Spencer, A.; Luís, A. (2012) *Clitics: An Introduction*. Cambridge: Cambridge University Press.
- Tobler, A. (1875). Review of Joules Le Coultre, *de l'ordre des mot dan Crestien de Troyes*. Extrait du programme de Pâques 1875 du Collège Vitzthum. Dresde 1875 ... [sic]. *Göttingische gelehrte Anzeigen*, 2(34), 1057-82. (Reprinted in his *Vermischte Beiträge zur französischen Grammatik*, vol. 5. Leipzig: Hirzel, 395-415).
- Tolstaya, M.N. Толстая М.Н. (2012). "Karpatoukrainskie enklitiki v južnoslavjanskoj perspective" Карпатоукраїнські енклітики в южнославянської перспективі (Carpathian Unkrainan Enclitics in a South Slavic Perspective). Карпато-балканський діалектний ландшафт; Язык и культура: 2009-2011 *Karpato-balkaniskij dialektnyj landshaft; Yazyk i kul'tura: 2009-2011* (Carpathian-Balkanian Dialectal Landscape: Language and Culture: 2009-2011), vol. 2. Moscow: Institute of Slavic Studies, 190-211.
- Vaillant, A. (1948). *Manuel du Vieux Slave*. Paris: Institut d'Études Slaves.
- Wackernagel, J. (1892). "Über ein Gesetz der indogermanischen Wortstellung". *Indogermanische Forschungen*, 1, 333-436.

- Wanner, D. (1996). "Second Position in Medieval Romance". Halpern, A.L.; Zwicky, A.M. (eds), *Approaching Second: Second Position Clitics and Related Phenomena*. Stanford: CSLI Publications, 537-78.
- Zaliznjak, A.A. Зализняк, А.А (1993). "Lingvistečeskie issledovanija" Лингвистические исследования (Linguistic Studies). Janin, V.L. Янин, В.Л.; Zaliznjak, A.A. Зализняк, А.А. (eds), *Novgorodskie gramoty na bereste. Iz raskopok 1984-1989* Новгородские грамоты на бересте (из раскопок 1984-1989 гг.) (Novgorod Birch Bark Letters from the Excavations 1984-1989). Moscow: Nauka, 191-344.
- Zalizniak, A.A. Зализняк, А.А. (2008). *Drevnerusskie enklitiki* Древнерусские энклитики (Old Russian Enclitics). Moscow: Yazyki Slavyanskoj Kultury.
- Zimmerling, A. (2012). "A Unified Analysis of Clitic Clusters in the World's Languages". *Computational Linguistics and Intellectual Technologies*, 11(18), 726-38.
- Zimmerling, A. Циммерлинг, А. (2013). *Sistemy porjadka slov slavyanskix yazykov v tipologičeskom aspekte* Системы порядка слов славянских языков в типологическом аспекте (Slavic Word Order Systems in Typological Perspective). Moscow: Yazyki slavyanskoj kultury.
- Zimmerling, A. (2015). "Parametrizing Verb-Second Languages and Clitic-Second Languages". *WORLDCOMP' 15. Proceedings of the 2015th International Conference on Artificial Intelligence*, vol. 1. CSREA Press, 281-7.
- Zimmerling, A.; Kosta, P. (2013). "Slavic Clitics: A Typology". *STUF*, 66(2), 178-214.
- Zwicky, A.M. (1977). *On Clitics*. Bloomington: Indiana University Linguistics Club.