

‘While’ and ‘Until’ Clauses and Expletive Negation in a Corpus of Bulgarian and Ukrainian Parallel Texts

Ivan Derzhanski
Institute of Mathematics
and Informatics,
Bulgarian Academy
of Sciences
iad58g@gmail.com

Olena Siruk
Institute of Philology,
Taras Shevchenko
National University of Kyiv;
IMI—BAS
olebosi@gmail.com

Abstract

The combination of the meanings ‘while’ and ‘until’ in a single lexeme and the use of expletive negation with the latter meaning are widespread phenomena that are a rich source of research problems. In this paper we present a comparative bilingual Bulgarian and Ukrainian corpus-based study of several conjunctions that share these two meanings. We discuss the difference in the frequency of expletive negation in the two languages, the use of *až* ‘even, all the way’ in Ukrainian and the impact of the original language in translated texts.

1. Introduction

The combination of the meanings ‘while’ and ‘until’ in a single lexeme and the obligatory or optional use of expletive negation¹ with the latter meaning are widespread phenomena (found in the Slavic languages, Hindi, Hungarian, Italian, Ket, Persian, etc.) that are a rich source of research problems, due to the high level of crosslinguistic and diachronic variation and the complex interaction of a variety of criteria, which makes it hard to obtain unambiguous judgements.

This paper presents a comparative bilingual Bulgarian and Ukrainian corpus-based study of several conjunctions that share these two meanings, with focus on the use of expletive negation.

The working Bulgarian–Ukrainian parallel corpus is composed entirely of fiction, including both original Bulgarian and Ukrainian texts and translations from other languages. The current version, which contains 100 texts (mostly novels, but occasionally parts of large novels, as well as collections of short stories by the same author and, if translated, by the same translator), is made of ten sectors, each composed of texts with the same original language and measuring approximately 800,000 words on the Bulgarian and 700,000 words on the Ukrainian side. This amounts to an approximate total of 15 million words in the entire corpus. Two sectors contain translations from Russian and two from English (because of the larger amount of material available); the remaining original languages are Bulgarian, French, German, Italian, Polish and Ukrainian. All texts are aligned at sentence level.

2. The Experiment

The lexical items studied in this experiment were conjunctions with the meaning ‘while, until’, especially such as allow expletive negation when in the latter meaning. In Bulgarian these terms are *dokato*, *dokle*, *dokogato* and *do(r)de(to)*.² In Ukrainian they are *doky*, *dopoky*, *zaky*, *pokil*’ and *poky*; in addition, the frequent combinations *až doky*, *až poky* ‘all the way until’ were treated as separate items, as was the particle *až* ‘even, all the way’ when it functions as a conjunction all by itself. All

¹ Also called pleonastic or paratactic negation, as opposed to semantic negation.

² We use the 1898 scientific transliteration system that is predominant in international linguistic publications on Cyrillic-written Slavic languages (Transliteration, 1898) for both Bulgarian and Ukrainian.

pairs of sentences or sentence fragments containing one of these items on one or both sides were located and counted.³ A total of 8809 such pairs were found in the corpus, including:

- 3446 pairs of *bona fide* matches,
- 3873 occurrences of unmatched Bulgarian ‘while/until’ items; among them are 843 that feature adverbial participles on the Ukrainian side⁴, 549 the temporal conjunction *koly* ‘when’ and 234 one of the compound conjunctions *v/u toj čas jak*, *tymčasom jak* and *todi jak* ‘while, whereas’,
- 1406 occurrences of unmatched Ukrainian ‘while/until’ items; among them 282 employ the conjunction *predi da* ‘before’ on the Bulgarian side and 202 have a verb of waiting in the matrix clause with an ‘until’ clause in Ukrainian and a *da*-clause in Bulgarian,
- and 84 pairs of sentences in which both sides contain one of the studied items, but the meaning is substantially different.

Among the Bulgarian items *dokato* dominates absolutely (6852 occurrences, or 92.56%). A distant second is *dodeto* (510 times, or 6.89%), which only appears in 29 of the 100 texts, with varying frequency, and only outnumbers *dokato* in the writing of one author (Bogomil Rainov); the only translations where *dodeto* has a tangible presence are JRR Tolkien’s *Narn i Hîn Hûrin*, translated by Lyubomir Nikolov, and Stendhal’s *Red and Black*, translated by Atanas Dalchev.⁵ Among the Ukrainian items the most common one is *poky* (3597 occurrences, or 72.87%), followed by *až poky* (587), *doky* (572), and *až doky* (103); (*až*) *doky* outnumbers (*až*) *poky* in only 7 texts out of 100.⁶

3. Semantics, Polarity and Aspectuality

The ‘while/until’ words in both languages combine several related meanings, which correlate, albeit not perfectly, with the aspectuality of the eventualities in both clauses and the polarity of the subordinate clause. Telicity, in turn, correlates to a high degree with Slavic aspect: imperfective and perfective verbs usually denote atelic and telic predicates, respectively.

		affirmative subordinate clause (Q)	
		atelic	telic
main clause (P)	atelic	P is happening <i>while</i> Q is happening	P is happening <i>until</i> Q happens
	telic	P happens <i>while</i> Q is happening	P happens <i>by the time (before)</i> Q happens

Table 1: The impact of aspect with an affirmative subordinate clause

A common exception to the correlation between aspect and telicity is a present tense form of an imperfective verb used for a telic event in vivid narration (1). Another is a perfective verb denoting its aftermath state (2):

(1) Bg: [...] *i otnovo vārvim, dodeto si davam smetka, če tova ne e nikakva ulica, a njakakāv pust mežduselski pāt.*

Uk: [...] *i os’ my znovu jdemo, poky ja usvidomljuju, ščo ce zovsim ne vulycja, a jakyjs’ bezljudnyj sil’s’kyj putivec’.*

³ Except where the term has a discernibly different meaning, as is the case with Ukrainian *doky* ‘until when? how long?’ and *poky* ‘for now, for the time being’. Contrariwise, the atemporal sense ‘whereas’ of Bulgarian *dokato* is hard to separate from the temporal one, so such occurrences were retained.

⁴ The frequency of this translation correspondence is discussed by Martinova-Ivanova (2015).

⁵ Notably, however, *dodeto* only appears twice in Nikolov’s translation of *The Hobbit* by JRR Tolkien and not at all in Dalchev’s translation of *The Gold Chain* by Alexander Grin, which demonstrates that such lexical preferences need not be a mark of the translator’s personal style, but instead may reflect his approach to the genre of the particular text, along with the fact that *dodeto* has come to be perceived as somewhat archaic.

⁶ The most pronounced preference for (*až*) *doky* is in Vasyl Zemliak’s *Green Mills* (71 occurrences, as opposed to only 4 of (*až*) *poky*); in *The Swan Flock* by the same author, however, we find (*až*) *poky* 43 times and (*až*) *doky* 39 times.

‘[...] and we’re on our way again, till I realise that this is not a street at all, but some deserted country road.’⁷

(Bogomil Rainov, *The Day Doesn’t Look Like the Morning*)

(2) Bg: *Ami toj šte izleze navān, dokato se sābličaš...*

Uk: *Ta vin vyjde, poky ty rozdjahatymešja...*

‘Well, he’ll go out [and stay outside] while you’re undressing ...’

(Pavel Vezhinov, *Libra*)

When the subordinate clause is negative, a third aspectual category comes into consideration, viz. consequent states of events, expressed in Bulgarian by perfect or pluperfect tense forms of perfective verbs but behaving as atelic predicates. In Ukrainian, which has replaced the aorist by the perfect, they assume the same form as telic predicates, although they can often be identified by the adverb *ščē* ‘still, yet’.

(3) Bg: *Iskam da ti obadja nešto, Džo, dokato ogānjat ne e ugasnal.*

Uk: *Ja xotiv by ščos’ tobi skazaty, Džo, poky šče vohon’ ne zhas.*

“Before the fire goes out [*in the translations*: while the fire hasn’t gone out (yet)], Joe, I should like to tell you something.”

(Charles Dickens, *Great Expectations*)

		negative subordinate clause (Q)		
		atelic	telic	perfect
main clause (P)	atelic	P is happening <i>while</i> Q is not happening	P is happening <i>until</i> Q happens (expletive negation)	P is happening <i>while</i> Q has not happened
	telic	P happens <i>while</i> Q is not happening		P happens <i>while</i> Q has not happened

Table 2: The impact of aspect with a negative subordinate clause

Negation in the subordinate clause is semantic if the predicate is atelic or perfect. With a telic predicate, as a rule, the negation is expletive. Exceptions, when the failure of a scheduled or recurrent event to happen is considered an event in its own right, are rare and potentially ambiguous.

(4) Bg: *Taka životāt prodālžavaše da si teče, dokato edin den kām sredata na januari srebristoto konče i ezdačkata mu ne se javiha na ugovorenata srešta.*

Uk: *Tak vono use tryvalo doty, doky odnoho dnja v seredyni sičnja sribljasto-čala kobylka ta jiji veršnycja v domovlenyj čas ne z’javyljsja.*

“So matters went on, till one day in the middle of January the silver-roan palfrey and its rider were missing [*in the translations*: did not show up] at the tryst.”

(John Galsworthy, *The Forsyte Saga*)

The semantic kinship between ‘*while* Q is happening’ and ‘*until* Q happens’ on the one hand, and ‘*until* Q happens’ and ‘*while* Q has not happened’ on the other, provides a rationale for the optional use of the expletive negation (Barentsen, 1979).

The construction may further imply that P terminates either *no sooner* than Q or *no later* than Q. The former meaning may be emphasised by Bg *pone*, Uk *prynajmni*, *xoč(a)* ‘at least’ and Bg *čak*, Uk *až* ‘all the way’; the latter, by Bg *samo*, Uk *lyše*, *til’ky* ‘only’. Expletive negation also indicates

⁷ English glosses are given in single quotes if they are ours, and in double quotes if they come from originals or published translations.

that P terminates *no sooner* than Q, but does so less strongly, and therefore can co-occur with the ‘only’ adverbs, although there no examples of this in our corpus.⁸

4. The Results

The corresponding constructions where a ‘while/until’ word is used in both languages are shown in Table 3. The rows correspond to the Bulgarian side and the columns to the Ukrainian one; ∂ stands for any ‘while/until’ word (except for Ukrainian *až*, which is represented and counted separately), \rightarrow for an atelic predicate, \odot for a telic one and $\odot \rightarrow$ for a resultative form (in Bulgarian only). The cell in the lower right corner (except for the totals) sums up nine occasions on which a sentence is interrupted after a ‘while/until’ word.

	$\partial \rightarrow$	$\partial \odot$	$až \partial \odot$	$až \odot$	$až \partial ne \odot$	$\partial ne \odot$	$\partial ne \rightarrow$	$\partial \dots$	Σ
$\partial \rightarrow$	1285	18	4			1			1308
$\partial \odot$	47	588	228	35	83	318			1299
$\partial ne \odot$	1	36	76	9	133	449	1		705
$\partial ne \odot \rightarrow$		1				76	4		81
$\partial ne \rightarrow$						4	40		44
$\partial \dots$								9	9
Σ	1333	643	308	44	216	848	45	9	3446

Table 3: Correspondences between ‘while/until’ structures

Sentences with ‘while’ clauses account for 40% of all. Bulgarian ‘until’ clauses correspond to Ukrainian ‘while’ clauses twice more often than the other way around.

The table makes it evident that Bulgarian uses expletive negation in ‘until’ clauses more sparingly than Ukrainian does (705 versus 216+848=1064 times). Where the Ukrainian does not employ it, the conjunction is twice more likely to be *až doky/poky* than simply *doky/poky* if there is expletive negation in the Bulgarian, but far less likely otherwise, which confirms the notion that the functions of *až* and expletive negation are related.

Among the 588 pairs of sentences with ‘until’ clauses in which there is no expletive negation in either language and no *až* in Ukrainian, there are 95 sentence pairs which state that P manages to happen by the time Q does (5) and 55 in which P measures the time until Q happens (6).

- (5) Bg: *A dokato se vārnete, šte si pogovorim za nešto seriozno s mis Džejn.*

Uk: *Poky vy povernetes’, ja dam mis Džejn dejaki nastanovy.*

“I’ll give Miss Jane a lecture till you come back.”

(Charlotte Brontë, *Jane Eyre*)

- (6) Bg: *Imame petnajset minuti, dokato Doktorāt prebroi negovite krāvni telca.*

Uk: *U nas je xvylyn p’jatnadcjat’ času, poky Likar poličyt’ joho erytrocyty.*

“We have fifteen minutes while [in the translations: until] the Doctor counts corpuscles.”

(Stanisław Lem, *Eden*; English by Marc E. Heine)

⁸ Examples from other sources: *Ti kaza, če edin loš šof’or e v bezopasnost samo dokato ne sreštnje drug loš šof’or, nali?* “You said a bad driver was only safe until she met another bad driver?” (F. Scott Fitzgerald, *The Great Gatsby*, translated by Neli Dospevska); *Tezi razsāždenija vārvtat mnogo gladko i strojno, no samo dokato njakomu ne hrumne da zapita: „A otkāde vsāštnost se e vzela tozi Praatom?”* “All this can be calculated very accurately and handsomely, but only until someone gets the idea of asking: “And just where did this Proto-Atom come from?”” (Stanisław Lem, *The Star Diaries*, translated by Lina Vasileva).

This group also includes all sentences in which the matrix clause is a polar question; the ‘no sooner’ meaning conveyed by expletive negation and by Ukrainian *až* appears incompatible with interrogation (Derzhanski, 1999).

- (7) Bg: *Bi li se sãgласil da vzemeš čantata mi, dodeto minem prez mitnicata?*

Uk: *Čy ne pohodyšsja ty vzjaty moju sumku, poky my projdemo mytnycju?*

‘Would you agree to take my bag until we go through customs?’

(Bogomil Rainov, *There Is Nothing Better than Bad Weather*)

If duration is stated, these devices are seldom used.

- (8) Bg: *Toj tičal cjala nošt, dokato stignal do pešterata.*

Uk: *I vin bih usju nič, poky distavsja do pečery.*

‘So he ran all the night till he came to the cave;’

(Rudyard Kipling, *The Jungle Book*)

- (9) Bg: *A posle prodālži i tuk, ne čas i ne dva, a celi tri dni, dodeto ne naučiha i majčinoto mi mljako.*

Uk: *A potim tryvav i tut, ta ne hodynu i ne dvi, a cilyx try dni, až poky mene ne vypatraly do ostann’oji kryxty.*

‘And then it [sc. the interrogation] continued here as well, not for an hour or two but for three whole days, until they got to know all my ins and outs.’

(Bogomil Rainov, *The Great Boredom*)

- (10) Bg: *Tazi borba prodālži polovin minuta, dokato cveteto ne se predade i ne uvisna bezžizneno v rākata na Pavliš.*

Uk: *Cja borot’ba tryvala pivxvylyny, poky kvitka ne zдалasja i neruxomo povysla v ruci Pavlyša.*

‘This struggle lasted for half a minute, until the flower gave up and hung limply in Pavlysh’s hand.’

(Kir Bulychev, *Village*)

Furthermore, expletive negation is hardly used in Bulgarian with a verb of waiting as the matrix predicate; there are only six examples of this, all of them in translations from other Slavic languages. In Ukrainian it appears 18 times.

- (11) Bg: *Šte se pribere otnovo v svoja Lubni i šte čaka mirno, dokato pronizitelnite trābi na Gradiv ne go prizovat otnovo kām podvizi...*

Uk: *Osjade u svojix Lubnax i čekatyme tyxo, až poky pronyzlyvi surmy Hradyvusa znov poklyčut’ joho...*

‘He will settle again in Lubni, and will wait quietly till the terrible trumpets call him to action again.’

(Henrik Sienkiewicz, *With Fire and Sword*; English by Jeremiah Curtin)

- (12) Bg: *[...] i tāj kato po mosta vārvjal goljam kervan natovareni muleta i kone, naložilo im se da počakat, dokato kervanāt se iztoči.*

Uk: *[...] ale po tomu mostu proxodyv same velykyj karavan nav”jučenyx muliv ta konej, i jim dovelos’ počekaty, poky vsi vony na toj bik ne perexopljat’sja.*

‘[...] and a caravan of pack-mules and sumpter-horses being in act to pass, it behoved them tarry till such time as these should be crossed over.’

(Boccaccio, *The Decameron*; English by John Payne)

In many cases, however, expletive negation has little impact in sentences which state that P lasts until Q. Because of this, sentences that denote very similar situations often differ only in its presence or absence.

- (13) Bg: *Svāržete me vednaga s Kiril Andreev, zvānete, dokato otgovori!*

Uk: *Z”jednajte mene zaraz z Kyrylom Andrejevym, dzvonit’, doky ne vidpovist’.*

‘Put me through to Kiril Andreev, ring until he answers!’

(Pavel Vezhinov, *Traces Remain*)

(14) Bg: *I gi izpālñjavaj, dokato toj ne te naznači za istinski gotvač.*

Uk: *I vykonuj, poky vin postavyt' tebe spravžnim axči.*

‘And fulfil them [*sc.* the commissions of the administrator of the Sultan’s kitchens] until he appoints you a regular cook.’

(Pavlo Zahrebelnyi, *Roksolana*)

Finally, in both languages there is an interrelation between the polarities of the matrix and the ‘until’ clause, so that negative matrix clauses show a strong preference for expletive negation.

		subordinate clause			Σ
		∂ ☀	∂ ne ☀	∂ ne ☀→	
main clause	affirmative	1267	409	72	1748
	negative	32	296	9	337
Σ		1299	705	81	2085

Table 4: The interplay of polarity in Bulgarian

		subordinate clause					Σ
		∂ ☀	až ∂ ☀	až ☀	až ∂ ne ☀	∂ ne ☀	
main clause	affirmative	619	277	41	173	600	1710
	negative	24	30	4	43	248	349
Σ		643	307	45	216	848	2059

Table 5: The interplay of polarity in Ukrainian

5. Variation by Source

The frequent optionality of expletive negation makes it a likely mark of the author’s or translator’s personal style and an area of influence of the original language.

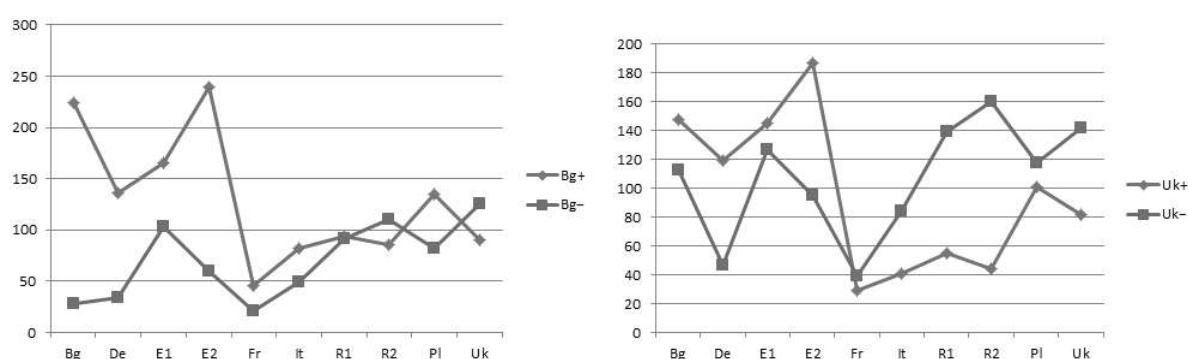


Figure 1: Quantities of affirmative and negative ‘until’ clauses by sector

Figure 1 shows the numbers of affirmative and negative ‘until’ clauses (without and with expletive negation) for each sector of the corpus in Bulgarian (on the left) and in Ukrainian (on the right). One can see that in the Bulgarian, German and both English sectors the affirmative ‘until’ clauses outnumber the negative ones in both languages, with the French sector coming close to this too. On the other hand, in the second Russian and the Ukrainian sector the negative ‘until’ clauses outnumber the affirmative ones in both languages, with the first Russian sector coming close.

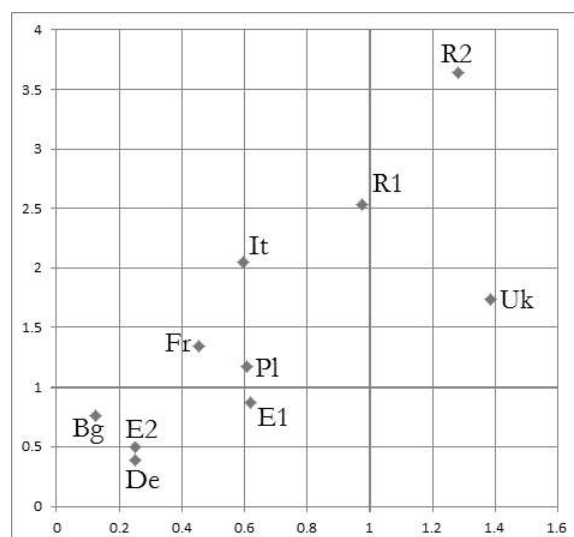


Figure 2: Ratio of negative to affirmative ‘until’ clauses by sector

Figure 2 shows the ratio of ‘until’ clauses with expletive negation to ‘until’ clauses without it for each sector of the corpus in Bulgarian (the x axis) and in Ukrainian (the y axis). It can be seen that on the Bulgarian side the lowest ratio of expletive negations is in the Bulgarian originals (which is to say that all translators use expletive negation more actively than the authors do) and the highest is in the translations from Ukrainian. On the Ukrainian side the lowest value is in the German sector. There is little doubt that the high frequency of expletive negation in ‘until’ clauses in Russian is the reason for which the Russian sectors of our corpus feature it in large quantities, but the distance between them, as well as between the two English sectors, proves that the individual authors and translators’ choices also play a significant part. Interestingly, the translations from Italian and Polish, the other two languages which use expletive negation in ‘until’ clauses, assume a middle position in the picture.

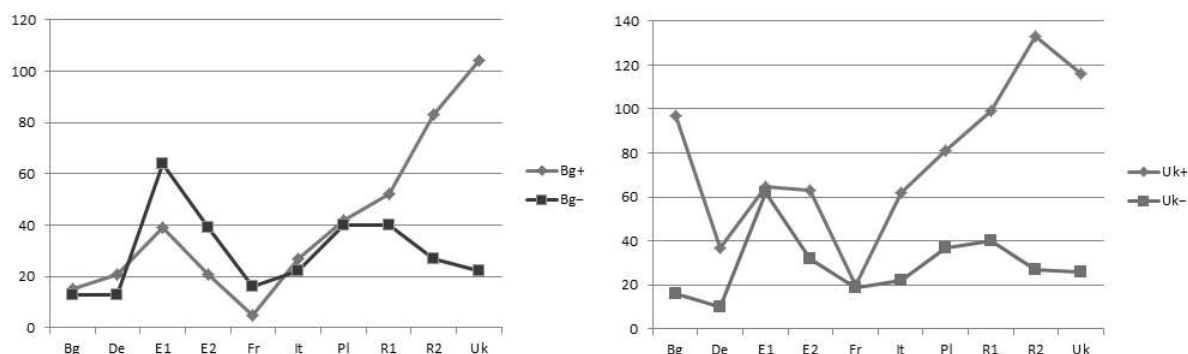


Figure 3: Quantities of affirmative and negative matrix clauses with a negative ‘until’ clause by sector

Figure 3 shows the numbers of affirmative and negative matrix clauses with a negative ‘until’ clause (with expletive negation) for each sector of the corpus in Bulgarian (on the left) and in Ukrainian (on the right). It is noteworthy that on the Bulgarian side in the French and both English sectors a negative ‘until’ clause is more often used with a negative than an affirmative matrix clause (reflecting the high frequency of ‘P won’t happen *until* Q’ constructions in these languages), which is never the case on the Ukrainian side, though the French sector and one of the English come very close.

6. Conclusions

Although Bulgarian and Ukrainian are closely related and share the key phenomena of this study (verbal aspect, lexical merging of ‘while’ and ‘until’, expletive negation with the latter meaning), the comparison reveals considerable differences. Expletive negation is much more frequent in Ukrainian than in Bulgarian, where it is more used in translations than in original writing and is largely a mark of the author’s style. In Ukrainian an important part is played by the particle and the conjunction *až*, which have no direct counterpart in Bulgarian (the particle is far more often used than Bulgarian *čak*, and the conjunction is not compatible with negation in the embedded clause). Finally, in translated text the grammar and usage patterns of the original language can have a significant impact on the translator’s choices. A more detailed study on a larger corpus could endeavour to look for possible diachrony effects as well.

References

- Barentsen, A. (1979). Nabljudenija nad funkcionirovanjem sojuza *poka*. In *Dutch Contributions to the Eighth International Congress of Slavists*. Lisse, 57–159.
- Derzhanski, I. (1999). Monotonicity and Interrogation. In Corblin, F., Dobrovie-Sorin, C., and Marandin, J.-M. (eds.), *Empirical Issues in Formal Syntax and Semantics 2: Selected Papers from the Colloque de Syntaxe et Sémantique de Paris (CSSP 97)*. The Hague: Thesus, 277–292.
- Martinova-Ivanova, P. (2015). Prevodăt na ukrainskite segašni deepričastija na bălgarski ezik. *Săpostavitelno ezikoznanie*, 2015(2):63–70.
- Transliteration (1898).
http://en.wikipedia.org/wiki/Scientific_transliteration_of_Cyrillic.

Proceedings of the
Second International Conference
Computational Linguistics in Bulgaria

**Computational
Linguistics in
Bulgaria**



9 September 2016
Sofia, Bulgaria

Table of Contents

Duško Vitas, Ljubomir Popović, Cvetana Krstev and Anđelka Zečević <i>How to Differentiate the Closely Related Standard Languages?</i>	1
Ivan Derzhanski and Olena Siruk <i>'While' and 'Until' Clauses and Expletive Negation in a Corpus of Bulgarian and Ukrainian Parallel Texts</i>	11
Verginica Barbu Mititelu and Elena Irimia <i>Linguistic Data Retrievable from a Tree-bank</i>	19
Ivelina Stoyanova, Svetlozara Leseva and Maria Todorova <i>Towards the Automatic Identification of Light Verb Constructions in Bulgarian</i>	28
Daša Farkaš, Matea Filko and Marko Tadić <i>HR4EU – Using Language Resources in Computer Aided Language Learning</i>	38
Atanas Atanasov <i>SynTags – Web Interface for Syntactic and Semantic Annotation</i>	47
Tsvetomila Mihaylova, Ivan Koychev, Preslav Nakov and Ivelina Nikolova <i>Finding Good Answers in Online Forums: Community Question Answering for Bulgarian</i>	54
Svetla Koeva, Ivelina Stoyanova and Martin Yalamov <i>Quotation Retrieval System for Bulgarian Media Content</i>	64
Bistra Popovska and Rositsa Dekova <i>Stress Patterns of Compounds and MWEs in English and Bulgarian</i>	74
Krešimir Šojat, Matea Filko and Daša Farkaš <i>Verbal Multiword Expressions in Croatian</i>	78
Sardar Jaf <i>A Simple Approach to Unifying Ambiguously Encoded Kurdish Characters</i>	86
Todor Lazarov <i>A Possible Solution to the Problem of Machine Translation of Verb Forms from Bulgarian to English</i>	95