

Temporal Adverbs and Adverbial Expressions 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
olebosi@gmail.com

Abstract

This paper presents a comparative bilingual corpus-based study of the use of several frequent temporal adverbs and adverbial expressions ('always', 'sometimes', 'never' and their synonyms) in Bulgarian and Ukrainian. The Ukrainian items were selected with the aid of synonym dictionaries of words and of set expressions, the corpus was used to identify their most common Bulgarian counterparts, and the frequencies of the correspondences were compared and scrutinised for possibly informative regularities.

1. Introduction

Although corpus-based contrastive research is quickly gaining momentum today, it is still at a very early stage, because parallel corpora are as yet available for few language pairs and the methods of their processing have only started being developed. In this paper we make a contribution to this field by addressing a pair of languages, Bulgarian and Ukrainian, which have received little attention in this regard, and evolving an algorithm for comparative analysis of lexical and phraseological units from a chosen lexical and semantic field on the basis of corpus and dictionary data.

The paper presents a comparative bilingual corpus-based study of the use of several frequent time adverbs and adverbial expressions with the meanings 'always', 'sometimes' and 'never' in Bulgarian and Ukrainian.

The working Bulgarian–Ukrainian parallel corpus (Siruk, Deržans'kyj, 2013; Siruk, Derzhanski, 2013) is composed entirely of fiction (mainly novels, although some shorter works have been included as well), including both original Bulgarian and Ukrainian texts and translations from other languages. The overall word count is currently 6.35 million in Bulgarian and 5.58 million in Ukrainian.¹ All texts have been aligned at sentence level with Hunalign (Varga et al., 2005), with subsequent manual correction of alignment errors.

A number of Ukrainian adverbs and adverbial expressions with the meanings 'always', 'sometimes' and 'never' were selected at the first stage of the research by means of a Ukrainian synonym dictionary (Burjačok et al., 1999) and a dictionary of phraseological synonyms (Kolomijec', Rehuševs'kyj, 1998). They formed a lexical and semantic group with three subgroups, composed of 14, 21 and 16 units respectively, not counting variants.

The use of both a word dictionary of synonyms and a phraseological one is expedient in order to achieve a more complete coverage of the semantic field because it is not always easy to estimate the power of the semantic cohesion of adverbial expressions. The dynamics of the forming of the common meaning of a cliché's constituents is also seen in the variations in the orthography of many of the

¹ The difference is due both to the contrast between the syntactic characters of the Bulgarian and the Ukrainian language (analytic *versus* synthetic) and the prevailing tendencies of the translators of the two schools towards comprehensiveness and conciseness, respectively.

expressions in both languages, whose components are written separately at first, but later can become a hyphenated or solid word.

At the second stage of the work, the corpus was used to find (through regular expression search) instances of the Ukrainian temporal adverbs and adverbial expressions drawn from the dictionaries and to identify their Bulgarian translation correspondences. These correspondences were checked against a Bulgarian dictionary of synonyms (Nanov, Nanova, 2000) and of phraseological synonyms (Nanova, 2005). Then they served in turn to locate further Ukrainian translation correspondences with the goal of expanding the lexical and semantic group. Finally, the frequencies (more precisely, the numbers of occurrences) of the matches were compared and studied.

2. ‘always’

The lexical and semantic group ZAVŽDY² is defined in the Ukrainian synonym dictionary as ‘all the time or over a certain temporal segment—invariably’ and contains 14 items, not counting variants: *postijno*, *povsjakčas*, *vsjakčas* [*usjakčas*], *doviku*, *poviky*, *povik*, *povik-viky*, *povik-vikiv*, *vvik* [*uvik*] (colloq.), *zavše* (colloq.), *zavsidy* [*zavsihdy*] (colloq.), *vse* [*use*] (colloq.), *zajedno* (dial.), *skriz*’ (rare) (Burjačok et al., 1999, vol. 1: 511, s.v. ZAVŽDY). All items in the group are single-word adverbs. The dictionary of Ukrainian phraseological synonyms does not feature such a group at all.

In the texts that make up the parallel corpus *zavždy* dominates absolutely (3,697 occurrences), followed by *nazavždy* ‘for ever’ 366, *vično* ‘eternally’ 278, *na viky* or *navik(y)* 271, *zavše* 201 and *postijno* ‘constantly’ 131.³

The makeup of the corresponding lexical and semantic group in Bulgarian is very similar. It is led by *vinagi* (3,462 occurrences), followed by *zavinagi* ‘for ever’ 446, *večno* 392, *postojanno* ‘constantly’ 390, *vsjakoga* (a close synonym of *vinagi*) 263, *naveki* 143 and *neizmenno* ‘invariably’ 94. The entry in the dictionary of synonyms (Nanov, Nanova, 2000: 61, s.v. VINAGI) also lists the stylistically marked items *vsjakogaž* (folk), *vsegda* (bookish, obs.) and *šjavga* (dial.), which do not occur in our corpus.

The numbers of occurrences of the translation correspondences involving the two languages’ most common ‘always’ items are given in Table 1. As in all tables in this paper, the Bulgarian words label the rows and the Ukrainian ones the columns.

	<i>zavždy</i>	<i>zavše</i>	<i>postijno</i>	<i>vično</i>	<i>nazavždy</i>	<i>naviky</i>
<i>vinagi</i>	2,191	129	20	13	20	3
<i>vsjakoga</i>	156	4			2	1
<i>postojanno</i>	53	1	41	3	3	
<i>večno</i>	48	1	3	202	1	1
<i>zavinagi</i>	20	2		4	250	67
<i>naveki</i>	1	2		2	8	86

Table 1: Correspondences between items expressing the meaning ‘always’.

Of some interest here is the high frequency with which Bulgarian *zavinagi* corresponds to Ukrainian *naviky*, and Bulgarian *postojanno* and *večno* to Ukrainian *zavždy*.

(1) Uk: *Odyn raz vidmovyššja vid svobody, a todi naviky zabudeš, ščo to take*. ‘You’ll relinquish freedom once, and then you’ll forget what it is for good.’

Bg: *Otkážeš li se vednāž ot svobodata, zabravjaš zavinagi kakvo e tja*.

(Pavlo Zahrebelnyi, *Roksolana*)

(2) Uk: *Vona zavždy tak žorstoko mene obražala!* ‘She always wounded me so cruelly.’

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

http://en.wikipedia.org/wiki/Scientific_transliteration_of_Cyrillic

³ Strictly speaking, ‘for ever’ is a different semantic field, but it has a significant overlap with ‘always’, and Bulgarian words from one field often correspond to Ukrainian words from the other.

Bg: *Tja postojanno me zasjagaše žestoko.*

(Charlotte Brontë, *Jane Eyre*)

(3) Uk: — *I Flores, vidljud'ko, zavždy poxmuryj Flores zasmijavsja.* ‘And Flores, the anchorite, Flores the always frowning, smiled.’

Bg: — *i Flores, samoživecăt, večno namrăštenijat, mračnijat Flores se zasmja.*

(Alexander Belyaev, *The Shipwreck Island*)

Generally, adverbs derived from the noun *vik* are more numerous and more frequent in Ukrainian than derivations of its etymological counterpart *vek* are in Bulgarian, in line with the fact that in Ukrainian this noun has a wider range of meanings (‘life, lifetime’ as well as ‘century’ and ‘age, epoch’, which are shared by the Bulgarian word as well).

3. ‘sometimes’

For the semantic field ‘sometimes’ the Ukrainian dictionary of synonyms lists 19 adverbs and set expressions, not counting variants (Burjačok et al., 1999, vol. 1: 644, s.v. *INODI*). The group only includes items which allow bounding from above (‘not always’, ‘not often’), as demonstrated by their ability to co-occur with the restrictive modifiers *lyše* and *til'ky* ‘only’. The dictionary of phraseological synonyms (Kolomijec', Rehuševskyj 1998: 62) gives four set expressions (clichés), two of which (*čas vid času* and *vid času do času*, actually variants of the same) are also given in the synonym dictionary. The group contains both adverbs and adverbial expressions with various structures.

In the corpus only four of these appear with significant frequencies, namely *inodi*, *časom*, *čas vid času* (with variants *čas od času*, *čas do času* and *vid/od času do času*) and *inkoly*, with 860, 828, 459 and 385 occurrences, respectively. They are followed by *raz u raz* (*raz po raz*, *raz za razom*) ‘time and again’ 317 and, far behind, by *zridka* 109, *vrjady-hody* (*urjady-hody*) 82 and *podekoly* 49.

In the matching Bulgarian sentences two items dominate: these are *ponjakoga* (with its rare variant *ponjavga*) and *ot vreme na vreme* (also written *otvreme-navreme*) ‘from time to time’, with 1,836 and 761 occurrences. Next come the pointedly colloquial *segiz-togiz* 59, *čas po čas* 51 and the archaic *navremeni* 47. The idiom *ot dăžd na vjatar* (lit. ‘from rain to wind’) only occurs seven times in the corpus. No occurrences were found of *ponjakogaž*, *sporadično*, *izrjadko*, *čat-pat* and *napăti*, which are also listed in the dictionary of synonyms (Nanov, Nanova, 2000: 436, s.v. *NJAKOGA*).

The distribution of translation correspondences is given in Table 2.

	<i>inodi</i>	<i>inkoly</i>	<i>časom</i>	<i>čas vid času</i>	<i>podekoly</i>	<i>zridka</i>	<i>vrjady-hody</i>	<i>raz u raz</i>
<i>ponjakoga</i>	603	255	452	43	35	13	10	8
<i>ot vreme na vreme</i>	55	28	102	295	12	34	35	40
<i>navremeni</i>	3	4	14	6		2	1	2
<i>segiz-togiz</i>	4	4	8	18		6	5	1
<i>čas po čas</i>	1		6		1			18

Table 2: Correspondences between items expressing the meaning ‘sometimes’.

It is obvious that the three frequent Ukrainian adverbs *inodi*, *inkoly* and *časom* are very similar in behaviour, and indeed the choice between them seems to be largely a matter of individual preference: there are texts in the corpus which use almost exclusively *inodi*, or nearly nothing but *časom*, or all three to an approximately equal extent. We may note, however, that *časom* corresponds to Bulgarian *navremeni* more often than the others, which may be accidental (given the shallow amount of data), though the correlation with the fact that both adverbs are derived from the nouns meaning ‘time’ (*čas* and *vreme*, respectively) is certainly interesting.

Another difference, concerning the co-occurrence of the adverbs and adverbial expressions with Bulgarian *samo* and Ukrainian *lyše* and *til'ky* ‘only’, is shown in Table 3.

	total	with 'only'	percentage
<i>ponjakoga</i>	1,835	28	1.53%
<i>ot vreme na vreme</i>	761	38	4.99%
<i>navremeni</i>	46	3	6.52%
<i>segiz-togiz</i>	59	9	15.25%
<i>inodi</i>	860	24	2.79%
<i>inkoly</i>	385	12	3.12%
<i>časom</i>	828	5	0.60%
<i>čas vid času</i>	459	10	2.18%
<i>zridka</i>	109	35	32.11%
<i>vrjady-hody</i>	82	14	17.07%
<i>podekoly</i>	49	0	0.00%

Table 3: Co-occurrence of some 'sometimes' items with 'only'.

If *inodi*, *inkoly* and *časom* are counted together, we see a strong correlation between them and the adverb *ponjakoga* on one hand, and between the set expression *čas vid času* (with its variants) and its near-literal counterpart *ot vreme na vreme* 'from time to time', on the other. The translators' tendency to stay close to the originals should explain this to some extent, but not entirely. A further contrast is shown in Table 4: the single-word adverbs are the only 'sometimes' items that often correspond to 'often' in the other language. This may indicate imprecise translation on some occasions, but the frequency with which it happens is too great to overlook, and suggests a semantic reason as well (greater proximity to the upper end of the frequency scale).

	<i>inodi + inkoly</i> + <i>časom</i>	<i>čas vid času</i>	<i>raz u raz</i>	<i>zridka</i>	<i>často</i> 'often'	<i>ridko</i> 'seldom'
<i>ponjakoga</i>	1,310	43	8	13	19	1
<i>ot vreme na vreme</i>	185	295	40	34	1	
<i>često</i> 'often'	70	9	18			
<i>rjadko</i> 'seldom'	4	1		21		

Table 4: Correspondences between 'sometimes' and 'often' or 'seldom' items.

One can note that Bulgarian *ot vreme na vreme* often corresponds to the structurally similar Ukrainian *raz u raz*.

The existence in Ukrainian of the adverb *zridka* (related to *ridko* 'seldom'), preferred host of *lyše* and *til'ky* 'only' and frequent translation correspondence of Bulgarian *rjadko* but with no precise counterpart in Bulgarian, constitutes yet another major difference between the two systems of expressions that lexicalise the meaning 'sometimes'.

(4) Uk: [...] *moja Kateryna tak varyt' galušky, ščo j het'manovi zridka dovodyt'sja jisty taki*. 'My Kateryna cooks such dumplings that even the hetman seldom gets to eat the like.'

Bg: *Mojata Katerina pravi takiva galuški, kakvito i hetmanăt rjadko može da jade*.
(Nikolai Gogol, *A Terrible Vengeance*)

Concerning the adjacent semantic field of 'sometimes' with no upper bound (items absent from the entry s.v. *INODI* in the Ukrainian synonym dictionary), the most conspicuous observations from the corpus are the frequent use of Ukrainian *raz u raz* and its variants and the high frequency of the iterative verb *buvaty* 'be occasionally, be regularly, happen' as a main verb or a parenthetic word, which corresponds to the Bulgarian adverb *ponjakoga* on 72 occasions:

(5) Uk: "*Istoryčna misija*", — *kazav, buvalo, Brjans'kyj*... "An historical mission," Bryansky used to say.'

Bg: “Istoričeska misija” — kazvaše *ponjakoga* Brjanski...
(Oles Honchar, *Guide-on Bearers*)

The Bulgarian verb *slučvam se* ‘happen’ has a similar function, but a much lower frequency; it corresponds to a Ukrainian adverb (*inodi, časom, inkoly, vryady-hody*) only 37 times.

(6) Uk: *Lyše inodi vin zryvajet’sja na kil’ka hodyn i raptom padaje, mov jastrub, pronyzanyj striloju.* ‘Only sometimes it starts up for several hours and suddenly falls down like a hawk pierced by an arrow.’

Bg: *Slučva se da duha i samo njakolko časa i izvednaž sekva, kato orel, pronizan ot strela.*
(Bolesław Prus, *Pharaoh*)

4. ‘never’

The lexical and semantic group *NIKOLY*, defined in both source dictionaries as ‘at no time, under no circumstances’, consists of nine adverbs, mostly stylistically marked ones, in the synonym dictionary (*zrodu* emph., colloq., *zrodu-viku* [*zrodu-zviku*] emph., colloq.; *doviku* emph., *povik* emph., *poviky* rare, *vik* emph., colloq., *povik-viky* [*povik-vikiv*] emph., poet., *vvik* [*uvik*] emph., colloq., *vovik* [*voviky*] arch., emph., colloq. (Burjačok et al. 1999, vol. 1: 1021, s.v. *NIKOLY*) and seven set expressions, not counting variants, in the dictionary of phraseological synonyms (Kolomijec’, Rehuševskij, 1998: 82). The data from the two dictionaries don’t intersect. The group is large in size, and its elements vary in structure. Among the three groups, this is the only one to contain set expressions with a high level of semantic cohesion, and two of these expressions were found in the parallel corpus, both times with different but likewise idiomatic Bulgarian translation counterparts:

(7) Uk: *Nu, to pobačyš joho, jak svoje vuxo.* ‘Well, you’ll see him as [you’ll see] your ear.’

Bg: *Šte go vidiš, kogato si vidiš vrata.* ‘... when you see your neck.’

(Henryk Sienkiewicz, *The Teutonic Knights*)

(8) Uk: *Jak rak svysne?* ‘When pigs fly?’, lit. ‘When the crayfish whistles?’

Bg: *Na kukovo ljato?* ditto, lit. ‘At cuckoo’s summer?’

(Bogomil Raynov, *Typhoons with Tender Names*)

In Bulgarian between 30 and 60 set expressions with the meaning ‘never’, not counting variants, are registered (Ničeva et al., 1974; Nanova, 2005):

(9) Uk: *Ajakže, čorta puxloho dočekaješja!...* ‘Oh sure, the hell you’ll live to see it!...’

Bg: *Kak ne, na kukovden!...* ‘Sure thing, on the first of Never!...’, lit. ‘on Cuckoo’s day’.

(Mykhailo Kotsiubynsky, *Fata Morgana*)

That said, the corpus in fact does little justice to the wealth of set expressions for ‘never’ that exists in either language. It does, however, feature some of the Ukrainian adverbs, especially *zrodu* (*zrodu-viku, zrodu-zviku*) with 187 occurrences, *doviku* with 94 and *povik* (*povik-viku*) with 33 (recall that the latter two have also the meaning ‘always’ or ‘for ever’ when used in affirmative contexts). Contrary to the synonym dictionary’s explicit statement, *zrodu* proves not to be limited to the past; it is applicable to the future as well:

(10) Uk: *Koly rozpovidaješ jim pro svoho novoho pryjatelja, vony zrodu ne pocikavljat’sja najistotnišym.* ‘When you’re telling them about your new friend, they will never be interested in the most substantial.’

Bg: *Kogato im razpravjate za njakoj nov prijatel, te nikoga ne vi pitat za naj-sašttestvenoto.*

(Antoine de Saint-Exupéry, *The Little Prince*)

While unusual in the use of *zrodu* for the future, this example is typical in that the Bulgarian uses the regular adverb *nikoga* ‘never’. The evidence of the corpus shows that Bulgarian has no other ‘never’ item comparable to Ukrainian *zrodu* in frequency, and the derivatives of *vek* are used with a negative meaning less often (and in the corpus not at all) than their Ukrainian etymological counterparts.

5. Conclusions

The comparative analysis of the lexical and semantic field of temporal adverbs and adverbial expressions on the basis of parallel texts makes it evident that this field is richer in synonyms in Ukrainian, whereas in Bulgarian it is generally more monolithic (this conclusion seems to be in variance with dictionary data,

but this can be explained with the fact that dictionaries cover specific ranges of genres, not restricted to fiction). The comparison of the frequencies with which words and expressions of one language correspond to words and expressions of the other in parallel sentences reveals subtle semantic oppositions and demonstrates the structure of the semantic fields and the relations between them.

As a side result of the search in the bilingual corpus, some items not marked in dictionaries as rare are shown to be so, which raises the question of checking the actual frequency of their use by the help of larger (and balanced) monolingual corpora.

The method employed in this investigation, which is readily applicable to other temporal adverbs and adverbial expressions and to other semantic fields, contributes to the comparative study of different languages' pictures of the world and, on a more practical level, holds potential for the improvement of synonym and bilingual dictionaries.

References

- Burjačok et al., 1999: Бурячок, А.А., Гнатюк, Г. М. (ред.) (1999). *Словник синонімів української мови*. Наукова думка, Київ.
- Kolomijec', Rehuševskuj, 1998: Коломиец Н. Ф., Регушевский Е. С. (1998). *Словник фразеологічних синонімів*. Ред. В.О. Винник. Издательство Радянська школа, Київ.
- Nanov, Nanova, 2000: Нанов, Л., Нанова, А. (2000). *Български синонимен речник*. София: Хейзъл.
- Nanova, 2005: Nanova, A. (2005). *Фразеологичен синонимен речник на българския език*. София; Хейзъл.
- Ničeva et al., 1974: Ничева, К., Спасова-Михайлова, С. Чолакова, Кр. (1974). *Фразеологичен речник на българския език*. Изд. на БАН: София.
- Siruk, Deržans'kuj, 2013: Сірук, О., Держанський, І. (2013), Лексичні перекладні еквіваленти в болгарських і українських паралельних текстах. *Українське мовознавство*, 43(2013):75–86.
- Siruk, O., Derzhanski, I. (2013). Linguistic Corpora as International Cultural Heritage: The Corpus of Bulgarian and Ukrainian Parallel Texts. *Digital Presentation and Preservation of Cultural and Scientific Heritage* (III/2013):91–98.
- Varga, D., Németh, L., Halácsy, P., Kornai, A., Trón, V., Nagy, V. (2005). Parallel Corpora for Medium Density Languages. In *Proceedings of the RANLP'2005*, pages 590–596. The program Hunalign: <http://mokk.bme.hu/resources/hunalign/>