Towards the Creation of a Belarusian Grammatical Dictionary

Ihor V. Shevchenko
ULIF NANU, Kyiv, Ukraine
Natalia Kotsyba
Warsaw University, Poland
Kiryl Kurshuk
Hrodna University, Belarus

Abstract
The article describes the process of creation of the Belarusian grammatical word-inflexion dictionary on the basis of linguistic similarities with the existing Ukrainian grammatical dictionary. The grammatical dictionary is developed to enable morphological mark-up of Belarusian texts and is the first tool of the kind for this language. The notions of the word-inflexion parameter and the word-inflexion class are introduced and explained. Cases of similarity and discrepancy in word-inflexion classes of the closely related languages are considered.

Key words: grammatical dictionary (GD), word-inflexion class (WIC), part of speech (POS), highly inflected languages, Slavic languages, Belarusian grammatical dictionary (BGD), Ukrainian grammatical dictionary (UGD).

1. Introduction

Grammatical dictionaries that provide description of the word-declination and word-formation are of great importance for inflective languages. Creation of an exhaustive set of lexemes' variants, as well as rules of word formation, belongs to most significant tasks in natural language processing as they enable lemmatization of word forms, i.e. their identification with the initial form available in dictionaries; morphological analysis and synthesis, grammatical tagging of text corpora.

2. Prototypes and the theoretical background

A grammatical dictionary, as we understand it, deals with word declination and has to contain all forms of inflected words of a corresponding language with their grammatical features. A great variety of such forms in Slavic languages makes such a task far from trivial. Talking about grammatical dictionaries we should first of all mention the fundamental work by Andrey Zalizniak [Zaliznyak 1967] that was the first thorough and systematic attempt to present a uniform and exhaustive description of the word-declination in a Slavic language. Even though the dictionary was not designed for computer-aided processing (back in the 1960's), it appeared to be quite applicable for it when computers became a bread-and-butter tool for linguists.

The Grammatical dictionary of the Ukrainian language (UGD) developed in the ULIF NANU [Shevchenko et al. 2005] provides by now a division of the lexemes fixed in dictionaries into 2456 word-inflexion grammatical classes (WIC), each of them presenting a set of lexeme endings according to their grammatical meanings, unique and uniform inside a class and therefore contains all the types of word-inflexion in Ukrainian [Shevchenko 1996]. The UGD has been one of the main tools supporting the first integrated Ukrainian lexicographical system

"Dictionaries of Ukraine" that has already seen eight CD-ROM editions (2001-2008) [Shirokov et al. 2007].

The uniformity of word-inflexion inside a WIC means, in terms of the computer processing of written texts, that all the words belonging to a WIC have the same set of the grammatical meanings and in each of the grammatical meanings (as well as their variants, if any) the same number of characters counting from the right is replaced with the same line of characters. Thus, words belonging to the same WIC differ in their invariable parts only.

As a matter of fact, a WIC is a set of words with the same type of word-inflexion, which is characterized by a set of values of the word-inflexion parameters [Shevchenko 1996, 2008]. The conformity of word-inflexion for many lexemes that are different in meaning and form allows to specify the following grammatical word-inflexion parameters.

1. Part of speech (or as its word-inflexion generalization – word-inflexion type)
2. Type of word stem
3. Conjugation pattern
4. Type of changes
5. Paradigm incompleteness
6. Aspect (for verbs)
7. Reflexivity (for verbs)
8. Imperative form (for verbs)
9. Passive participle suffix
10. Gender (for nouns)
11. Denotatum type (for nouns)
12. Form of the genitive case for masculine nouns
13. Form of the locative for masculine and neutral nouns
14. Form of the dative for masculine nouns
15. Form of the accusative case in plural (for nouns)
16. Atypical word-form features in certain grammatical meanings
17. Type of the accent distribution in the word-inflexion paradigm.

A WIC is determined by a combination of the parameter values, for example, all masculine nouns of the second declension that indicate human beings, end in a soft consonant, drop the vowel *-e-* in indirect cases and do not have atypical features in their inflexion belong to the same WIC (in our classification #1540), examples are: "v<u>y</u>borec'" ("voter"), "irl<u>a</u>ndec'" ("Irishman").

lexeme	part of speech	declension	basis	change	animacy	genitive		WIC
v<u>y</u>borec' **v<u>y</u>davec'** **promysl<u>o</u>vec'**	n	2dec	soft	-e	person	a		1540

lexeme	part of speech	conjugation	basis	final suffix	aspect	reflex	change	WIC
kompensuv<u>a</u>ty **likviduv<u>a</u>ty** **normalizuv<u>a</u>ty**	v	1dec	iota	-aty	imperf+ perf	–	–	382

lexeme	part of speech	declension	basis	change	animal coord	pecul	WIC	
Vitalijiv zmijiv nencjiv	a	possessive	hard	ji-je	+	–	2335	
vsjakyj dekotryj žodnyj	pron	general	hard	–	+	–	1145	

The initial form of a word recorded in a dictionary as a word entry is the object of identification. A word-inflexion parameter makes sense only for certain groups of words according to their grammatical function. For example, there is hardly any use considering the option "form of the genitive" and the "gender" for a verb (if we deal with gender in general as a characteristic feature of a lexeme); likewise, the parameter "aspect" or "form of imperative" with respect to a noun or an adjective. Thus, a parameter is characterized within its domain. The word-inflexion parameter can be regarded as a discrete function with a limited range of possible values (the value area). As an example, the well-known list of parts of speech can be given. The parameter "type of the word stem" can get one of 5 values: hard, soft, combined, iotacized and r-type. The parameter "gender" has potentially 10 different values for a lexeme (three genders, their six combinations by the order of two and, besides, one combination of all three genders), while the form of the genitive for masculine nouns has three values: *-a* (or *-ja*, depending on the word ending), *-u* (*-ju*), or both *-a* (*-ja*) and *-u* (*-ju*) are possible.

Each of the word-inflexion parameter values can be implemented only for words that possess a certain form. For example, Ukrainian verbs can end only in *-ty*, or *-tysja*. No word entry with a different form can be considered as a candidate for a verb in the process of grammatical identification[1]. At the same time, the word that ends in *-mu* is not necessarily a verb. For example, it can be a *pluralia tantum* noun, for example: "graty" ("bars"). Thus, the words with the ending *-ty* and *-tysja* set a domain for the parameter. Belonging to a domain does not imply that the parameter value of the word is really implemented. In such cases it is logical to talk about an optional parameter since a word that ends in *-ty* can be but not necessarily is a verb. There are, however, some cases when a word is determined by a parameter value unconditionally. For example, any Ukrainian feminine noun whose last consonant is *-k* changes it into *-c* in the dative and locative cases. Thus, for this very productive group of words, the above value of the change is mandatory and one can talk about an area of mandatory implementation for the parameter value.

3. The material basis for the Belarusian Grammatical Dictionary

The problem of full and explicit descriptions of the word-inflexion exists for other Slavic languages as well. In most of them it has been already resolved one way or another. However, this cannot be said about the Belarusian language yet. The problem of the creation of the Belarusian grammatical dictionary is being solved by us with the help of an electronic version of the *Dictionary of the Belarusian Language. Spelling. Orthoepy. Accentuation. Word-Inflection.* ("Slounik belaruskaj movy. Arfahrafija. Arfaepija. Akcentuacyja. Slovazmjanenne") that deals first of all with the spelling and some elements of orthoepy [BelOrthDict 1987]. The dictionary covers the main stock of the Belarusian vocabulary in its state by the end of the 20th century covering the material from *The Belorussian-Russian Dictionary* (Moscow, 1962), *Explanatory Dictionary of the Belarusian Language* ("Tlumačal'ny slounik belaruskaj movy", vol. 1-5, 1977-

[1] Perhaps with the exception of expressions like *spatky, jiston'ky*, ("to sleep", "to eat") which are classified as predicatives in the UGD. However, they are treated as diminutive forms of the infinitive in the UGTag, see [Kotsyba, Mykulyak, Shevchenko 2009] in the same volume and [Derzhanski, Kotsyba 2008] for the discussion about predicatives.

1984), as well as the materials of the lexical card file of the Yakub Kolas Institute of Linguistics. The word-inflexion in the dictionary is given in an abridged form, typical for a spelling dictionary of Slavic languages, so that the non-initial word-forms are presented starting from the last letter common to all forms followed by other spelling forms in different grammatical meanings.

abvjazac' *zak.,* abvjažu, abvjažaš, -ža, -žam, -žace, -žuc'
kacic' *nezak.,* kaču, kociš, kocic', kocim, kocice, kocjac'
vyhan *s.-g.,* -nu, -ne, -nau; *(dzejanne)* -nu, -ne
nasenne [n'ne] -nni
svjakruxa -usje, -ux

The dictionary presents options for certain grammatical forms and meanings:

drukarnja -rni, -ran' *i* -rnjau
madel' -llu [l'lu], -lej *i* –ljau

The volume of a paradigm presentation varies depending on the rolling accent:

knot knota, -oce, *mn.* knaty, -tou
les lesu, lese, *mn.* ljasy, ljasou
svat svata, svace, *mn.* svaty, -tou
sklep -pa, -pe, *mn.* skljapy, -pou

The dictionary specifies the position of the accent for all the word forms.

Given all the information contained in the orthographic dictionary, it can be an invaluable source of grammatical information that can be used for morphological analysis. For this purpose, the content of the paper dictionary had to be brought to the electronic form.

4. Digitalization as a preparatory stage of creating the grammatical database

The paper dictionary containing over 900 pages with words presented in three columns at each page was scanned and OCRed with the help of the Fine Reader 8.0 program for optical character recognition. The specifics of the dictionary text and its layout demanded a careful and painstaking post-OCR correction. The insufficient quality of the dictionary text recognition, including improper splitting of the text into paragraphs, excessive bolding etc., presented a serious problem while dealing with its electronic form, making it impossible in many cases to automatically identify word entries' limits. One of the biggest problems was the lack of recognition sets for the characters of accentuated Cyrillic letters, which strongly distorts the original text and significantly increases the time for its correction.

Due to the inadequate for Belarusian texts recognition software and poor quality of scanned pages the scanning and text recognition of the dictionary text through relevant computer programs generated a lot of mistakes. During the correction of the dictionary text after its OCR we found some regular and irregular errors and inadequacies. Regular errors could be corrected by the AutoCorrect editing function in the MS Word, the rest of the errors were to be corrected manually, which required a lot of time and still did not guarantee a hundred-percent accuracy.

The most frequent regular text recognition errors include:

1. u (u under stress) passed in the letter ŭ (u semivowel): abahaču → abahačŭ.

This problem was partially solved by replacing through the AutoCorrect the combination of "a consonant + 'ŭ'" by "a consonant + 'y'".

2. o (o under stress) was misrecognized as follows: abloki → ablbki

To correct these errors we also used the AutoCorrect function of the MS Word: the combination of "a consonant + 'b' + a consonant" (this combination is almost never found in the

Belarusian language) was replaced by “a consonant + ‘o’ + a consonant”. However, here we needed to replace separately each combination of letters: “rbk”, “lbs”, “lbp” and many others. For example, when using the AutoCorrect function on the combination “rbm” to change it into „rom” the program returns about 250 substitutions. Here one should be very careful as the combination of the type “a consonant + b + a consonant” can be, even though only occasionally, found in the Belarusian language, e.g. in the word “adbrakavany” (“rejected”). That is why this procedure had to be controlled.

3. e (e under stress) in some cases was transformed into *ë*: smecce → smëcce.

4. In some cases the character „y” was misrepresented by the combination “’i”.

Apart from spelling errors, after the recognition the text lost its paragraphs’ layout, which made it problematic to separate word entries automatically.

An example of a dictionary text distorted by the recognition program:

ablbki -kau, *adz.* vbblaka, *častej* vbblaki

vbblaka -ku, *mn.* voblaki *i* ablbki, vbblakau *i* ablbkau

vbblaki -kau, *adz.* voblaka *i* ablbki

5. Grammatical forms and lexemes to be reconsidered

It should be noted that some of existing word-formation models that are present in the orthographic dictionary are regarded presently by Belarusian linguists as Russisms, and some corrections will have to be introduced into the BGD later, cf. also [Lomtev 1956, Pivtorak 1997].

For example, the suffix of the Russian origin *-cel’* is to be replaced by other Belarusian specific suffixes: dubicel’ → dubil’nik (“tanning agent”), akislicel’ → akislial’nik (“oxidizer”), zbavicel’ → zbauca (“saviour”), natxnicel’ → nathnial’nik (“inspirer”), vyxavacel’ → vyxaval’nik (“tutor”).

The adjective suffix *-jonn-*, considered unnatural for the Belarusian language, is being replaced with more relevant ones: pensijonny → pensijny (adj. “pension”), sesijonny → sesijny (adj. “session”) and the suffix *-ann-* with *-avy*: firmenny → firmavy (adj. “brand-name”), dareformenny → dareformavy (“pre-reform”).

Therefore, the dictionary published in 1987 contains some words with suffixes that are not recommended for use anymore.[2]

6. Mapping Belarusian WICs on the Ukrainian Grammatical Dictionary

The idea of creating a Belarusian grammatical dictionary is based on a wide parallelism in word-inflexion of the two closely related languages – Ukrainian and Belarusian. The proximity in lexical composition and an apparent parallelism in the word-inflexion systems of both languages, and in some cases non-flexion modifications (change proper, insertions, omissions) are easy to notice [Pivtorak 1997]. Thus, it is assumed that nouns of neutral gender of *-nne*: “stajanne” (“standing”), “abohatvarenne” (“idolizing”), “abyhravanne” (“playing up”), and others are similar in their inflexion to the Ukrainian WIC 2108 covering *singularia tantum* nouns of neutral gender ending with *-nnja*: “stojannja” (“standing”), “maljuvannja” (“drawing”), while those ending with *-sc’*: “lehal’nasc’” (“legality”) „lehkavažnasc’” (“light-mindedness”) and many others meet the Ukrainian WIC 2143 covering singular nouns of the 3rd declension with the change *o-i* in some cases: “aktyvnist’” (“activity”), “raptovist’” (“suddenness”). Adjectives ending with *-y* “bely” (“white”), “ahul’ny” (“general”) obviously have the Ukrainian corresponding WIC 2302 “bilyj” (“white”), “spil’nyj” (“common”) that brings together adjectives with the hard ending. In the verb ending with *-ac’* “dbac’” (“take care”), “spac’” (“sleep”) we recognize the Ukrainian WIC 697

[2] The problem is not crucial at the initial stage of our work. Most probably both russified and authentic forms will be necessary in the grammatical dictionary to correctly identify as many forms as possible that occur in texts. This is a sociolinguistic issue and needs further research.

“db<u>a</u>ty” (“take care”), “sp<u>a</u>ty” (“sleep”), i.e. verbs of the 1st conjugation with iotacized endings in the present tense and without passive participle in the paradigm. Similarly, the infinitive form ending in *-acca*: “abahaščacca” (“get rich”), “zžyvacca” (“get used”) roughly corresponds to the WIC 700 “vyhyn<u>u</u>tysja” (“bend”), “zupyn<u>y</u>tysja” (“stop”), and so on.

Language	Lexeme	part of speech	declension	basis	change	animacy	genitive	WIC
Ukr.	**stoj<u>a</u>nnja**	n	2dec	hard	–	person	a	2108
Bel.	**abahrav<u>a</u>nne**	n	2dec	hard	–	animate	a	2108
Ukr.	**rapt<u>o</u>vist’**	n	2dec	hard	i-o	inanimate	a	2143
Bel.	**leh<u>a</u>l’nasc’**	n	2dec	hard	–	inanimate	a	2143

Language	Lexeme	part of speech	declination	basis	change	animal coord	pecul	WIC
Ukr.	**bilyj**	adj	general	Hard	–	+	–	2302
Bel.	**ah<u>u</u>l’ny**	adj	general	Hard	–	+	–	2302

Language	Lexeme	part of speech	conjugation	basis	final suffix	aspect	reflex	change	WIC
Ukr.	**db<u>a</u>ty**	v	2dec	iota	-aty	imperf	–	–	697
Bel.	**dbac’**	v	2dec	iota	-ac’	imperf	–	–	697
Ukr.	**zupyn<u>ia</u>tysja**	v	2dec	iota	-jaty	imperf	+	–	700
Bel.	**abahaščacca**	v	2dec	iota	-acca	imperf	+	–	700

Such analogies allow us to make the first rough division of the Belarusian dictionary entries into specific word-inflexion proto-classes that require further differentiation. We can ascribe WIC numbers to a large part of lexemes presented in the dictionary through a series of global replacements based on the ending of forms, as well as elements of word-inflexion paradigms available in the dictionary. Further refinement of the word stock, as well as the detection of rare classes, is to be carried out manually.

7. Further differentiation of Belarusian WICs

We should note that the existence of far-going analogies between WICs of the Ukrainian and Belarusian languages does not mean a total coincidence of the endings or their word-inflexion parameters. For example, in Belarusian there is no vowel change in the feminine nouns ending in *-asc’*: “tv<u>o</u>rčasc’” (“creativity”), “pramysl<u>o</u>vasc’” (“industry”), which is inherent for similar Ukrainian nouns ending with *-ist’*: “vlučnist’ – vlučnosti’” (“marksmanship”) in a number of indirect cases.

Let us consider the process of differentiation of Belarusian word-inflexion classes. For example the WIC 1607 covers masculine nouns of the 2nd declension with stems ending in a hard consonant and *-a* flexion in the genitive without vowel change, designating inanimate objects, e.g., Ukr. “hryb” ("mushroom"). The closest counterpart to this Ukrainian WIC can be found in similar, vowel-invariable, Belarusian entries like "mar<u>o</u>z" (“frost”). At the same time, there are some types of change for this group of Belarusian nouns that are not inherent in the Ukrainian word-inflexion, for instance, the change *t - c* “abanem<u>e</u>nt” – loc. “abanem<u>e</u>nce” (“season ticket”), *d - dz* „pad’<u>je</u>zd” – loc. “pad’<u>je</u>zdze” (“doorway”), as well as the double change of the lexeme “sneh” (“snow”) with the locative “sn<u>e</u>ze” and the nominative plural “snjah<u>i</u>”. The WICs 1615, 1627 and 1635 respectively are differentiated on the basis of these lexemes. Similarly, the Ukrainian WIC 2134, represented by the lexeme “p<u>o</u>vist’” (“novel”), corresponds in the Belarusian word-inflexion system at least to two classes: the WIC 2134 “ap<u>o</u>vesc’” (“novel”)

and the WIC 2137 “kosc’” (“bone”), the latter having the change of the stem vowel in indirect cases *o-a* (kosc’ – kascj).

Language	lexeme	part of speech	Declension	basis	animacy	genitive	change	WIC
Ukr.	**hryb**	n	2dec	hard	inanimate	a	–	1607
Bel.	**marɔz**	n	2dec	hard	inanimate	a	–	1607
Bel.	**abanement**	n	2dec	hard	inanimate	a	t-c	1615
Bel.	**pad’jezd**	n	2dec	hard	inanimate	a	d-dz	1627
Bel.	**sneh**	n	2dec	hard	inanimate	a	h-z, je-ja	1635
Ukr.	**pɔvist’**	n	3dec	soft	inanimate	a	–	2134
Bel.	**apɔvesc’**	n	3dec	soft	inanimate	a	–	2134
Bel.	**kosc’**	n	3dec	soft	inanimate	a	o-a	2137

In some cases we have noticed phonetic phenomena that are not typical for the Ukrainian language like the appearance of the prothetic *v-* in some grammatical meanings. This also predetermines further differentiation of word-inflexion classes. For example, the Ukrainian WIC 1991, which combines nouns of neutral gender with hard endings, corresponds to several Belarusian WICs. In addition to the obvious WIC 1991 “haspadarstva” (“economy”), which reflects the main type of the word-inflexion in the neutral gender, we have also lexemes like “aknɔ” (“window”, WIC 2001) and “vɔzera” (“lake”, WIC 2002). In the first case we can observe the appearance of the inserted *v-* in the plural (aknɔ – vɔkny), in the second case it is the omission of the same *v-* (vɔzera – azjɔry).

languag e	Lexeme	part of speech	declen sion	basis	gender	anim	genitive	change	WIC
Ukr.	**haspadarstvo**	n	2dec	hard	n	inanim ate	a	–	1991
Bel.	**haspadarstva**	n	2dec	hard	n	inanim ate	a	–	1991
Bel.	**aknɔ**	n	2dec	hard	n	inanim ate	a	+v	2001
Bel.	**vɔzera**	n	2dec	hard	n	inanim ate	a	-v	2002

The Ukrainian verbal class 490 represented by "ohornuty" (“embrace”) in its basic version corresponds to the Belarusian WIC 490 represented by the lexeme “nedacjahnuc’” (“fail to hold out”), but also to the WIC 491, the inflexion paradigm with the change of *e-*: abamknuc’, -ne, -nješ, -nu, -njem, -njace, -nuc’ (“surround”), and the WIC 494 with a stem change *a-o* (abharnuc’, abharne, abhɔrneš (“embrace”).

language	lexeme	part of speech	conjugation	basis	final suffix	aspect	change	WIC
Ukr.	**ohornuty**	v	1conj	hard	uty	perf	–	490
Bel.	**nedacjahnuc’**	v	1conj	hard	uc’	perf	–	490
Bel.	**abamknuc’**	v	1conj	hard	uc’	perf	e-ë	491
Bel.	**abharnuc’**	v	1conj	hard	uc’	perf	a-o	494

Meanwhile, some Ukrainian word-inflexion parameters are not inherent in the Belarusian language and such WIC as 1628, whose peculiarity is in the change *k-č* in the vocative case: “čolovik” (“man”) – voc. “čoloviče” (“o, man!”), has no counterpart in the Belarusian language

since the modern literary language does not have this case at all. Therefore, the Belarusian counterparts of the Ukrainian lexemes from the WIC 1628 like “čalavek” (“a man”) correspond to the more general class of masculine nouns ending with a *-к*, which is the WIC 1788 represented by “mistyk” (mystic”), where the change in the vocative case is not observed.

Let us have a look at another example of the same phenomenon. In the Ukrainian language there are different types of inflexion for nouns designating people and those for animals because the accusative plural case of words designating animals is realized in two optional forms. One of them equals the nominative plural and the other coincides with the genitive: “pasty koni” and “pasty konej” (“to graze horses”). In the case of human beings, only the form coinciding with the genitive is acceptable: “zustrity divčat”, not “*zustrity divčata” (“to meet girls”). For the Belarusian language such a differentiation does not exist. Therefore, the Belarusian counterparts like “zajčyk” (“little hare”) or “vožyk”(“hedgehog”) of Ukrainian nouns from the WIC 1789 (“zajčyk”) and 1629 (“jižak”) belong in the Belarusian word-inflexion system to the same WIC 1788.

language	lexeme	part of speech	declension	basis	vocative change	animacy	genitive	WIC
Ukr.	**mistyk**	N	2dec	hard	–	person	a	1788
Bel.	**mistyk**	N	2dec	hard		animate	a	1788
Ukr.	**zajčyk**	N	2dec	hard	–	animal	a	1789
Bel.	**zajčyk**	N	2dec	hard		animate	a	1788
Ukr.	**čoloviк**	N	2dec	hard	k-č	person	a	1628
Bel.	**čalavek**	N	2dec	hard		animate	a	1788
Ukr.	**jižak**	N	2dec	hard	k-č	animal	a	1629
Bel.	**vožyk**	N	2dec	hard		animate	a	1788

In that case in fact we have a convergence of at least four different Ukrainian WICs in one Belarusian. Or, from another point of view, differentiation of one Belarusian inflexion class into several Ukrainian ones.

At the same time we can see here an opposite example where one Ukrainian inflexion class differentiates in the Belarusian word-inflexion: the Belarusian word “junak” (“young man”) corresponding to the Ukrainian word with the same spelling and meaning from the WIC 1788 falls into a different inflexion class (in our classification it is the WIC 1787) due to the changed ending *-ou* in the genitive plural: “junakou”, unlike the more conventional *-au* (čalavekau, vožykau). Such a differentiation is unknown in Ukrainian: “junakiv”, “čolovikiv”, “jižakiv”.

language	lexeme	part of speech	declension	basis	plural change	anim	genitive	WIC
Bel.	**čalavek**	n	2dec	hard	–	animate	a	1788
Bel.	**vožyk**	n	2dec	hard	–	animate	a	1788
Bel.	**junak**	n	2dec	hard	a-o	animate	a	1787

8. Conclusions and further work

Nowadays, the importance of common tools and formats in the NLP is accentuated by many scholars. This enables easier convergence, alignment and reusability of natural language resources. The mapping method of developing new ones presented here, especially for less resourced languages like Belarusian, can help to catch up with better-resourced languages in shorter time and with less effort. At the same time, the common format and conceptual apparatus is preserved and the existing tools and expertise are reused.

Controlled compiling of a dictionary described above is connected with discovering the grammar of a language in detail on a large (and, hopefully, exhaustive) amount of language material and can be compared to a corpus driven approach of grammar study. Ideally, it should be projected on existing grammar descriptions to serve as their complementation. Further check-up on large corpora data is necessary as well, of course.

The described analysis can also give an account about comparative morphology of Belarusian and Ukrainian and serve a material basis for the comparative grammar of those languages.

Apart from the additional advantages for language description brought by the mapping method of creating a grammatical dictionary, its direct usage cannot be overestimated.

To sum up, grammatical dictionaries present a number of advantages:

- GDs can be used in a variety of ways, e.g. the statistics of usage given by a GD can help us trace more common patterns of word-inflexion in similar classes of words, which can be useful for recommendations on standardization, considering the current variability of existing forms in both Ukrainian and Belarusian. Statistics of WICs can be of use in grammatical homonymy disambiguation.
- GDs can be a powerful tool for comparative studies too, a much neglected by computational linguistics area so far.
- GDs are corpus-driven, so they help us reveal the information about a language that is not covered in grammars, or is not covered consistently or clear enough for the users.

Thus, creating a Belarusian Grammatical Dictionary allows us, on the one hand, to obtain a detailed word-inflexion classification of the Belarusian language, and on the other hand, to study and summarize the differences in the word-inflexion systems of closely related languages. At the moment we have in our classification about 400 word-inflexion classes for Belarusian. The work on the grammatical dictionary of the Belarusian language is going on and will lead to the creation of a morphological analyzer and tagger for the Belarusian language.

Bibliography

1. BelOrthDict (1987). Slounik belaruskaj movy. Arfahrafija. Arfaepija. Akcentuacyja. Slovazmjanenne. Minsk: „Belaruskaja Savetskaja Encyklapedyja" imya P. Brouki.
2. Derzhanski I. and Kotsyba N. (2008). The category of predicatives in the light of the consistent morphosyntactic tagging of Slavic languages. In *Lexicographic Tools and Techniques: Proceedings of the MONDILEX First Open Workshop*, pages 68–79, Moscow: IITP—RAS.
3. Kotsyba N., A. Mykulyak and I. V. Shevchenko (2009). UGTag: morphological analyzer and tagger for Ukrainian language (in this volume).
4. Lomtev T. P. (1956). Grammatika belorusskogo jazyka. Moscow.
5. Pivtorak H. (1997). Bilorus'ka mova. Kyiv: Lybid;. 240 p.
6. Shevchenko I. V. (1996). Algorytmična slovozminna klasyfikacija ukrajins'koji leksyky. *Movoznavstvo.* Vol. 4–5, Kyiv, p. 40–44.
7. Shevchenko I. V., Shirokov V.A., Rabulets' A.G. (2005). Elektronnyj grammatičeskij slovar' ukrainskogo jazyka. In: Proceedings of the international conference „Megaling'2005. Prikladnaja lingvistika v poiske novyx putej", 27.VI–2.VII 2005. Meganom, Crimea, Ukraine, p. 124–129.
8. Shevchenko Ihor (2008). Parametryzacija jak osnova hramatyčnoji identyfikaciji slovnykovyx odynyc' ukrajins'koji movy. In: Prykladna lingwistyka ta lingvistyčni texnolohiji. Megaling-2007. Zbirnyk naukovyx prac'. Kyiv: Dovira, p. 393–402.
9. Shirokov V. A. (1998). Informacijna teoria leksykohrafičnyx system. Kyiv: Dovira, p. 16, 152–174, 274–288.
10. Shirokov V. A., Rabulets' O. H., Shevchenko I. V. , Kostyshyn O. M., Yakymenko K. M. (2007). Intehrovana leksykohrafična systema „Slovnyky Ukrajiny", version 3.1. Kyiv. CD-edition.
11. Zaliznyak A. A. (1967). Russkoje imennoje slovoizmenenije. Moscow: Nauka, 370 p.