

СЛОВАРНЫЕ ПРОЕКТЫ И ТРУДЫ

УДК 81'27

DOI: 10.17223/22274200/9/3

З.И. Резанова, Г.Ю. Веснина

ПОДКОРПУС РУССКОЙ РЕЧИ БИЛИНГВОВ ЛИНГВИСТИЧЕСКОГО КОРПУСА «ТОМСКИЙ РЕГИОНАЛЬНЫЙ ТЕКСТ»: ПРИНЦИПЫ РАЗМЕТКИ И МЕТАРАЗМЕТКИ КОРПУСА¹

В статье представлено описание подкорпуса речи русскоязычных билингвов лингвистического корпуса «Томский региональный текст»: характеризуются особенности материала (тип двуязычия), предопределяющие необходимость коррекции разметки и метаразметки подкорпуса по отношению к основному корпусу. Цель создания подкорпуса речи билингвов – выявление грамматической и семантической интерференции первого языка (L1) в рамках второго (русского – в данном случае) языка (L2). Для этого в разметку корпуса вводятся специальные теги, которые маркируют факты интерференции.

Ключевые слова: корпусная лингвистика, лингвистический корпус, лингвистическая разметка, региональная лингвистика, билингвизм, языковая интерференция, русский язык.

Данная статья является продолжением серии публикаций, посвящённых Томскому региональному корпусу текстов, разрабатываемому в Лаборатории когнитивных исследований языка филологического факультета Томского государственного университета. Ранее были описаны ключевые параметры построения корпуса: принцип региональности [1], сбалансированность и репрезентативность [2, 3]. В рамках настоящей публикации обсуждаем проблемы разметки и метаразметки подкорпуса русской речи билингвов.

Цель проекта по созданию корпуса «Томский региональный текст» – представить структуру общения на региональном варианте русского языка, характерном для Томска и Томской области. Корпус стремится к отображению полидискурсивности и к пропорциональному представлению языковой ситуации в регионе. Поскольку в настоящее время территорию Томской области

¹ Публикация подготовлена в рамках проектов РГНФ №14-14-70010 и РФФИ № 14-36-50404.

населяют представители более 120 национальностей (русские – 92,1%, татары – 1,7%, украинцы – 1,1%, немцы – 0,9%, чувашаи – 0,4%, узбеки – 0,4%, азербайджанцы – 0,4%, белорусы – 0,3% и др.) и функционирует более 40 национально-культурных объединений, включая 6 региональных и 15 местных национально-культурных автономий¹, невозможно не учесть би- и полилингвизм как компонент языковой ситуации региона. Таким образом, появление подкорпуса обусловлено стремлением отразить языковое своеобразие Томской области.

Подкорпус речи билингвов наследует некоторые черты учебных корпусов (*learner corpora*), эритажных корпусов (*heritage language corpora*) и корпусов ошибок (*error corpora*). Среди известных работ в этом направлении, на которые мы ориентировались, можно назвать корпус русского как иностранного и русского как эритажного языка RLC [7], международный корпус английского языка как иностранного ICLE [8], корпус немецкого языка Falko [9] и работы S. Granger [10, 11].

Специфика проектируемого нами подкорпуса считают в том, что в нём будет представлена речь людей, для которых русский язык не является первым, но уровень его освоения стремится к уровню владения первым языком. Кроме того, это не коллекция текстов, созданных в результате учебных заданий, а свободные беседы на определённые темы (работа, семья, язык).

1. Параметры метаразметки

Цель создания подкорпуса речи билингвов – формирование лингвистически размеченного собрания текстов, отражающих специфику русского языка в условиях его существования в билингвальных ситуациях, а также ряд социокультурных характеристик билингвов, потенциально способных повлиять на способы функционирования языка. Специфика русской речи билингвов находит отражение в фиксации в системе *разметки* подкорпуса фактов языковой интерференции, которая проявляется в ошибках использования языковых единиц разного типа. В *метаразметке* корпуса фиксируется информация об этно- и

¹ В исследовании в данном аспекте мы опирались на данные источников: Национальные объединения. Администрация Томской области, Официальный сайт Томского отделения «Росстата» [4, 5], см. также [6].

социокультурных характеристиках билингов – типе второго языка, степени овладения им, а также ряде социальных характеристик (см. далее). Соотнесение параметров разметки и метаразметки дает возможность пользователю подкорпуса делать целенаправленные выборки текстов, позволяющих соотнести тип языковой интерференции и социокультурные характеристики билинга.

Своеобразие билингвальных ситуаций заключается в языке контактирования, который обуславливает влияние как на структуру русского языка, так и на речевую реализацию этого влияния. Поэтому нас интересует тип языка, который является контактирующим в каждой конкретной ситуации. Как следствие, в метаразметке подкорпуса появляется два особых параметра, а именно: этническая принадлежность говорящего и языка, которыми он владеет. Так как этническая принадлежность определяется человеком в силу самосознания и отнесения себя к какому-либо этносу, значение этого параметра для каждого текста определяется: а) в аудиозаписях – прямым опросом информантов; б) в архивных документах и виртуальных текстах – исходя из анализа косвенных текстовых данных.

В исследованиях, посвящённых изучению параметров билингвизма, часто встаёт вопрос обозначения языковой пары. Существуют следующие пары терминов: родной – неродной (иностранной), первый – второй. Поскольку при попытке определить родной язык мы сталкиваемся с целым набором критериев, такими как первоочерёдность усвоения языка, принадлежность к определённому этносу, язык окружающего общества, уровень владения языком [12], из которых лишь первоочерёдность является неоспоримой, нам представляется разумным придерживаться пары терминов первый и второй языки, обозначив их, как это принято при исследовании билингвизма, соответственно L1 и L2. При этом язык исследуемого этноса может занимать любую из этих позиций.

Аналогично данным об этнической принадлежности информанта данные о L1 и L2 информантов выявляются для устных текстов методом прямого опроса, в отношении архивных документов – предположительно исходя из анализа косвенных текстовых данных. При этом по аналогии со снятой и неснятой омонимией пользователь корпуса получает возможность работать как с текстами, где точно

определены языковые пары, так и с текстами с неснятой неопределённостью.

При изучении процессов овладения L2 важны социальная среда, тип овладения языком – бытовой, естественный, в ходе контактов с носителями или же учебный, системный, способ овладения языком. Значимым является не только путь, но и степень овладения языком. Согласно выработанным Советом Европы общеевропейским компетенциям выделяют три широких уровня владения языком: А (элементарное владение), В (самостоятельное владение) и С (свободное владение) – и два подуровня в каждом из них – А1 (уровень выживания), А2 (предпороговый уровень), В1 (пороговый уровень), В2 (пороговый продвинутый уровень) С1 (уровень профессионального изучения), С2 (уровень владения в совершенстве) [13]. Уровень владения языком, несомненно, влияет на степень интерференции, однако на данном этапе, без проведения специализированных тестов, мы можем оценить уровень владения тем или иным языком для каждого информанта лишь относительно. Согласно Вайнрайху этот относительный уровень владения можно определить, установив, какой язык является доминирующим вследствие того, что одним из языков билингв владеет лучше [14].

По сравнению с метаразметкой, приводимой для корпуса ICLE [8], мы не включаем в рассмотрение языковую технику автора текста и условия обучения, однако оставляем прочие параметры. Метаразметка подкорпуса билингвов включает: данные о месте и времени записи текста, Ф.И.О. собирателя, данные об информанте (Ф.И.О., пол, год рождения, образование, социальное положение, национальность, используемые языки L1, L2 и относительный уровень владения ими), данные о тексте (форма, тип, дискурс, субжанр, жанр, тема) (табл. 1).

По сравнению с параметрами описания L1 и L2, приведёнными в корпусе немецкого языка Берлинского университета им. Гумбольдта Falko, мы не указываем время освоения L2 и языковое окружение, поскольку при рассмотрении письменных текстов начала XX в. эти данные точно установить невозможно.

Таблица 1. Пример метаразметки в подкорпусе

Информация о месте записи текста	Томск
Информация о времени записи текста	Ноябрь 2014
Собиратель	NNN
Информант	
• ФИО	NNN
• Пол	мужской
• Год рождения	1935
• Образование	среднее специальное
• Социальное положение	пенсионер
• Этническая принадлежность	татарин
• L1	татарский
• Уровень владения	доминирующий
• L2	русский
• Уровень владения	второстепенный
Форма текста	устный
Тип текста	диалог
Дискурс	личностный
Группа субжанров	информативные
Жанр	беседа
Тема	Семья (приезд внука)

2. Параметры разметки

Специфика разметки подкорпуса русской речи билингвов состоит в фиксации фактов языковой интерференции при помещении текстов в подкорпус. Так как интерференция охватывает все уровни языка – фонетический, морфологический, семантический, синтаксический, и размечаться она должна на каждом из этих уровней. Однако поскольку среди задач создания данного подкорпуса и корпуса в целом не значилась фиксация фонетических признаков, отображается только морфологическая, семантическая и синтаксическая интерференция.

При разработке параметров разметки мы принимали во внимание те параметры, которые выделяются создателями корпусов ошибок и параллельных корпусов переводных текстов [15, 16]. Опираясь на классификации ошибок, представленные в работах S. Granger [17], мы вводим категории, представленные в табл. 2.

В отличие от решений, предлагаемых S. Granger, пунктуационные ошибки и опечатки или описки в подкорпусе русской речи билингвов не размечаются, поскольку при транскрипции устных текстов данные параметры не являются релевантными.

Таблица 2. Типология ошибок и тэги разметки

Интерференционные ошибки по уровням языка		Интерференционные ошибки в конкретных категориях и процессах	
#ДЕР#	Деривационные ошибки	#ПРЕ# #СУФ#	Префиксация Суффиксация
#ГР#	Грамматические ошибки	#РОД# #ЧИС# #ПАД# #НКЛ# #ВИД# #ВИД# #ЗЛГ# #РЕФ#	Род Число Падеж Наклонение Вид Время Залог Рефлексивность
#ЛЕКС#	Лексические ошибки	#ЗНЧ# #СОЧ# #L1# #L2# #L1+L2#	Неверное значение Неверная лексическая сочетаемость Калькирование из первого языка Калькирование из второго языка Модифицированное заимствование, имеющее компоненты первого и второго языка
#СИН#	Синтаксические ошибки	#ПОР#	Неверный порядок слов

В морфологической разметке в подкорпусе, как и в основном корпусе, фиксируются следующие параметры: морфологические категории, лексико-грамматические категории. К этому мы добавляем фиксирование отклоняющихся грамматических категорий.

(1) кое-какое #ГР# #РОД# подвижка

(2) больше пять #ГР# #ПАД# лет

(3) Вот он у меня проучил #ГР# #РЕФ# там сентябрь, октябрь

На лексическом уровне в основном корпусе предусматривается только распределение слов по лексико-семантическим группам. В подкорпусе маркируем заимствования из L1 или L2. Следовательно, вводится помета заимствованных слов: #L1#, #L2#:

(4) Конечно / кызым #L1# / сейчас сложнее//

Кроме того, при помощи тэгов будут отмечаться отклонения от лексического значения слова при интерферентных ошибках в лексико-семантической сочетаемости, изменение формы слова:

(5) Какой он переводчик? Ну так/ для обзора #ЛЕКС# #ЗНЧ#.

(6) ну отпишитесь дусларчики #ЛЕКС# #L1+L2#!!!

(7) Однако, в большинстве своем, танцам в Татарстане характерны #ЛЕКС# #СОЧ# светлые и добрые замыслы #ЛЕКС# #ЗНЧ#.

Синтаксическая разметка в основном корпусе ориентирована на передачу дискурсивных особенностей речи в регионе и предполагает членение текста на формально-содержательные единицы (текст, блок, предложение) и характеристику риторических связей между ними (в соответствии с теорией риторической структуры У. Манна и С. Томпсон [18, 19]), обеспечивающих связность текста. В подкорпусе русской речи билингов целевая направленность обуславливает введение дополнительных компонентов разметки, связанных с проявлением в структуре текста интерференционных влияний. Несмотря на то, что в русском языке нет устойчивого порядка слов, при нарушении устойчивых конструкций, обусловленном синтаксическими особенностями иного языка, данные факты размечаются тэгами #СИН#, #ПОР#:

(8) *Тот короче путь.* #СИН# #ПОР#

Выводы

Итак, основным принципом разметки и метаразметки подкорпуса русской речи билингов корпуса «Томский региональный текст» является в первую очередь соответствие принципам основной разметки и метаразметки всего корпуса. В метаразметке данного подкорпуса учитываются параметры «национальность» и «L1», «L2» – языки, которыми владеет информант. Для каждого из языков указывается приблизительный уровень владения.

Для грамматической, семантической и синтаксической разметки подкорпуса русской речи билингов важным является учёт фактов интерференции на всех уровнях языка. С этой целью была разработана классификация ошибок, которые могут быть обусловлены интерференцией, выявляемых на каждом уровне языка.

На данный момент собрана начальная база текстов и ведётся её разметка. В ближайшей перспективе запланировано включение первичной разметки в программную оболочку (созданную под руководством В.В. Поддубного, см. [20]) и её тестирование. Апробация может привести к необходимости скорректировать изначальные параметры.

Литература

1. Мишанкина Н.А. Лингвистический корпус «Томский региональный текст»: теоретико-методологическое обоснование проекта // Вестн. Том. гос. ун-та. – 2014. – № 389. – С. 28–37.
2. Sologub Olga, Rezanova Zoya, Temnikova Irina. The Concept of the Tomsk Regional Corpus: Balance and Representativeness // The XXV annual international academic conference, Language and culture, 20–22 October 2014 / Procedia – Social and Behavioral Sciences, 154 (2014). – P. 175–178.
3. Резанова З.И. Лингвистический корпус «Томский региональный текст»: типологически релевантные параметры сбалансированности и репрезентативности // Вестн. Том. гос. ун-та. Филология. – 2015. – №1(33). – С. 38–50.
4. Национальные объединения. Администрация Томской области [Электронный ресурс]. – URL: <http://tomsk.gov.ru/ru/grazhdanskoe-obschestvo/natsionalnye-ob-edineniya> (дата обращения: 16.11.2014).
5. Официальный сайт Томского отделения “Росстата” [Электронный ресурс]. – URL: http://tmsk.gks.ru/wps/wcm/connect/rosstat_ts/tmsk/ru/statistics/ (дата обращения: 16.11.2014).
6. Резанова З.И. Дискурсивные стратегии презентации национально-культурной идентичности // Вестн. Том. гос. ун-та. Культурология и искусствоведение. – 2012. – № 4 (8). – С. 40–54.
7. The Russian Learner Corpus (RLC). URL: <http://web-corpora.net/RussianLearnerCorpus/search/> (дата обращения: 16.11.2014).
8. International Corpus of Learner English v2. – URL: <https://www.uclouvain.be/en-277586.html> (дата обращения: 16.11.2014).
9. Das Falko-Handbuch Korpusaufbau und Annotationen, Version 2.01 - Reznicek, Marc; Lüdeling, Anke; Krummes, Cedric und andere (2012). – URL: <http://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik/forschung/falk> (дата обращения: 16.11.2014).
10. Granger S. Corpus d'apprenants, annotation d'erreurs et ALAO: une synergie prometteuse // Cahiers de Lexicologie. – Vol. 91, no. 2. – P. 117–132. – URL: <http://hdl.handle.net/2078.1/75650> (дата обращения: 16.11.2014).
11. Granger S. The Learner Corpus: A Revolution in Applied Linguistics // English Today 39, Vol. 10, no. 3. – P. 25–29 (1994). – URL: <http://hdl.handle.net/2078/75699>. (дата обращения: 16.11.2014).
12. Чиршева Г.Н. Родной и неродной языки в условиях формирования раннего детского билингвизма // Филологический класс. – 2014. – №1 (35). – С. 101–104. – URL: <http://cyberleninka.ru/article/n/rodnoy-i-nerodnoy-yazyki-v-usloviyah-formirovaniya-rannego-detskogo-bilingvizma> (дата обращения: 17.11.2014).
13. Common European Framework of Reference for Languages Learning, Teaching, Assessment. – Council of Europe, Language Policy Unit, Strasbourg, 1986. – URL: <http://www.coe.int/lang-CEFR>
14. Вайнрайх У. Языковые контакты. – Киев: Вищ. шк., 1979. – 364 с.
15. Al-Jarf R. Spelling error corpora in EFL // US-China Foreign Language. Sino-US English Teaching. Vol. 7, No.1 (Serial No.73). – P. 6–15.

16. *Díaz-Negrillo Ana & Ángel García-Cumbreras M.* A tagging tool for error analysis on learner corpora International Computer Archive of Modern and Medieval English. – URL: <http://clu.uni.no/icame/ij31/ij31-page197-204.pdf>.

17. *Granger S.* Error-tagged Learner Corpora and CALL: A Promising Synergy // *CALLIC Journal*, 20 (3). – P. 465–480.

18. *Mann William, Matthiessen Christian, Thompson Sandra A.* Rhetorical structure theory: A Framework for the Analysis of Texts. Reprinted from *IPRA Papers in Pragmatics*. – Vol. 1. – Washington, 1987. – 22 p.

19. *Литвиненко А.О.* Описание структуры дискурса в рамках теории риторической структуры: применение на русском материале // *Труды Междунар. семинара Диалог'2001 по компьютерной лингвистике и ее приложениям*. – Аксаково, 2001. – С. 159–168.

20. *Поликарпов А.А., Поддубный В.В., Кукушкина О.В., Кубарев А.И., Варламов А.А., Суровцева Е.В., Пирятинская Е.Ф.* Комплексная тексто-аналитическая система «СтилеАнализатор-2», основанная на Web-технологиях: разработка, наполнение данными и тестирование на прикладных задачах // *Сайт лаборатории общей и компьютерной лексикологии и лексикографии МГУ*, 2014, <http://istina.msu.ru/publications/article/5848839/>

META-DATA AND ANNOTATION DESIGN OF THE RUSSIAN-SPEAKING BILINGUALS SPEECH SUBCORPUS IN THE STRUCTURE OF THE TOMSK REGIONAL CORPUS

Voprosy leksikografii – Russian Journal of Lexicography. 2016. 1 (9). 29–39.

DOI: 10.17223/22274200/9/3

Rezanova Zoya I., Vesnina Galina Yu., Tomsk State University (Tomsk, Russian Federation). E-mail: resso@rambler.ru / resso@mail.tsu.ru / galina.y.vesnina@gmail.com

Keywords: corpus linguistics, linguistic corpora, markup, regional linguistics, bilingualism, language interference, Russian.

This article continues a series of publications devoted to the Tomsk Regional Corpus of texts developed by the Laboratory for Cognitive Studies of Language of the Faculty of Philology, Tomsk State University. The purpose of the project of the Tomsk Regional Corpus is to present the structure of communication in the regional version of the Russian language typical of Tomsk and Tomsk Oblast. The corpus tends to display the polydiscursivity and the proportional representation of the linguistic situation in the region. The subcorpus of bilinguals' speech inherits some features of learner corpora, heritage language corpora and error corpora. The article describes the Russian-speaking bilinguals speech subcorpus of the Tomsk Regional Text corpus: features of the material (type of bilingualism) that determine the need for markup and meta-markup correction of the subcorpus in relation to the main corpus are characterized. The meta-markup of the bilinguals' subcorpus includes: data on the place and time of text recording, the collector's name, data about the informant (name, gender, year of birth, education, social status, nationality, languages used (L1, L2) and their level), data about the text (form, type, discourse, subgenre, genre, topic). The purpose of the bilinguals speech subcorpus is to identify the grammatical and semantic interference of the first language (L1) with the second (Russian in our case) language (L2). Thus, the morphological markup of the subcorpus, like that of the main corpus, registers the following parameters: morphological

categories, lexical and grammatical categories. The authors also register the deviating grammatical categories. The syntactic markup in the main corpus is focused on the reflection of the discursive features of speech in the region, and involves division of the text into formal and substantive units (text, block, sentence) and description of rhetorical links between them that ensure coherence of the text. The purpose of the subcorpus of the Russian speech of bilinguals requires the introduction of additional markup components connected with the interference effects that show in the structure of the text. The corpus markup uses special tags that mark the facts of interference.

References

1. Mishankina, N.A. (2014) Linguistic corpus “Tomsk regional text”: theoretical and methodological background of the project. *Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal*. 389. pp. 28–37. (In Russian).
2. Sologub, O., Rezanova, Z. & Temnikova, I. (2014) The Concept of the Tomsk Regional Corpus: Balance and Representativeness. *Procedia – Social and Behavioral Sciences*. 154. pp. 175–178.
3. Rezanova, Z.I. (2015) Tomsk Regional Corpus: typologically relevant parameters of balance and representativeness. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 1(33). pp. 38–50. (In Russian). DOI: 10.17223/19986645/33/4
4. Tomsk Oblast Administration. (n.d.) *Natsional'nye ob'edineniya* [National associations]. [Online]. Available from: <http://tomsk.gov.ru/ru/grazhdanskoe-obschestvo/natsionalnye-ob-edineniya>. (Accessed 16 November 2014).
5. The official website of the Tomsk branch of Rosstat. [Online]. Available from: http://tmsk.gks.ru/wps/wcm/connect/rosstat_ts/tmsk/ru/statistics/. (Accessed 16 November 2014).
6. Rezanova, Z.I. (2012) Discourse strategies of presentation of national cultural identity. *Vestnik Tomskogo gosudarstvennogo universiteta. Kul'turologiya i iskusstvovedenie. – Tomsk State University Journal of Cultural Studies and Art History*. 4 (8). pp. 40–54. (In Russian).
7. The Russian Learner Corpus. [Online]. Available from: <http://web-corpora.net/RussianLearnerCorpus/search/>. (Accessed 16 November 2014).
8. International Corpus of Learner English v2. [Online]. Available from: <https://www.uclouvain.be/en-277586.html>. (Accessed 16 November 2014).
9. Reznicek, M. et al. (2012) *Das Falko-Handbuch Korpusaufbau und Annotationen, Version 2.01* [The Falko manual corpus construction and annotation, version 2.01]. [Online]. Available from: <http://www.linguistik.hu-berlin.de/institut/professuren/korpus-linguistik/forschung/falk>. (Accessed 16 November 2014).
10. Granger, S. (2007) Corpus d'apprenants, annotation d'erreurs et ALAO: une synergie prometteuse [Learner corpora, annotation errors and CALL: a promising synergy]. *Cahiers de Lexicologie*. 91:2. pp. 117–132. [Online]. Available from: <http://hdl.handle.net/2078.1/75650>. (Accessed 16 November 2014).
11. Granger, S. (1994) The Learner Corpus: A Revolution in Applied Linguistics. *English Today*. 39:10:3. pp. 25–29. [Online]. Available from: <http://hdl.handle.net/2078/75699>. (Accessed 16 November 2014).

12. Chirsheva, G.N. (2014) Native and non-native languages in the development of early childhood bilingualism. *Filologicheskij klass*. 1 (35). pp. 101–104. [Online]. Available from: <http://cyberleninka.ru/article/n/rodnoy-i-nerodnoy-yazyki-v-usloviyah-for-mirovaniya-rannego-detskogo-bilingvizma>. (Accessed: 17 November 2014). (In Russian).

13. Council of Europe, Language Policy Unit. (1986) Common European Framework of Reference for Languages Learning, Teaching, Assessment. Strasbourg. [Online]. Available from: <http://www.coe.int/lang-CEFR>.

14. Weinreich, U. (1979) *Yazykovye kontakty* [Language contacts]. Kiev. Vishcha shkola.

15. Al-Jarf, R. (2010) Spelling error corpora in EFL. *US-China Foreign Language. Sino-US English Teaching*. 7:1 (Serial No. 73). pp. 6–15.

16. Díaz-Negrillo, A. & García-Cumbreras, M.A. (2007) A tagging tool for error analysis on learner corpora. *ICAME Journal*. 31. pp. 197–204. [Online]. Available from: <http://clu.uni.no/icame/ij31/ij31-page197-204.pdf>.

17. Granger, S. (2003) Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO Journal*. 20 (3). pp. 465–480.

18. Mann, W., Matthiessen, C. & Thompson, S.A. (1987) *Rhetorical structure theory: A Framework for the Analysis of Texts*. Reprinted from IPRA Papers in Pragmatics. Vol. 1. Washington.

19. Litvinenko, A.O. (2001) [Description of the structure of discourse within the rhetorical structure theory: application to Russian material]. *Trudy Mezhdunarodnogo seminara Dialog'2001 po komp'yuternoy lingvistike i ee prilozheniyam* [Proceedings of the International Workshop Dialog'2001 on computational linguistics and its applications]. Aksakovo. pp. 159–168. (In Russian).

20. Polikarpov, A.A. et al. (2014) *Kompleksnaya teksto-analiticheskaya sistema "StileAnalizator-2", osnovannaya na Web-tehnologiyakh: razrabotka, napolnenie dannymi i testirovanie na prikladnykh zadachakh* [Integrated text-analytical system "StileAnalizator-2" based on Web-technology: the development, filling and testing of data on applied problems], [Online]. Available from: <http://istina.msu.ru/publications/article/5848839/>.