



Исследование функционирования эквивалентов слова в тексте на материале Украинского национального лингвистического корпуса

*Широков В.А., доктор технических наук, академик
Лучик А.А., доктор филологических наук, профессор
Остапова И.В., кандидат технических наук
Яблочков Н.М., аспирант*

В статье рассматриваются критерии разграничения эквивалентов слова и фразеологизмов. Показано, что установление частотности будет способствовать более глубокому пониманию таких единиц. Исследование предполагается провести на базе Украинского национального лингвистического корпуса, который представляет собой унифицированный, структурированный, размеченный и лингвистически компетентный массив текстов на естественном языке.

• Украинский национальный лингвистический корпус • эквивалент слова • фразеологизм • частотность

The article deals with the criteria of word equivalents delineation as well as phraseology, it is proved that their frequency establishment will assist deeper understanding of such units. In order to establish the correctness of calculations it is offered to make a research by the help of the Ukrainian National Linguistic Corpus that was unified, structured, tagged and considered to be philological competent array of natural language text.

*• Ukrainian National Linguistic Corpus • the word equivalent
• phraseological unit • frequency*

Лингвистический термин «эквивалент слова» и его понимание, предложенное Г. П. Рогожниковой, как устойчивого сочетания, характеризующегося «целостностью значения, преимущественно постоянной, неизменной формой», в речевом потоке имеющего «обычно одно словесное ударение» [10, с. 4], сегодня распространён и с некоторыми дополнениями активно функционирует в украинской языковедческой науке. Подтверждением сказанному может служить издание собственно украинских [6], а также переводных русско-украинского и украинско-русского, украинско-польского и польско-украинского словарей эквивалентов слова [7]. На необходимость вынесения эквивалентов слова в отдельные подрубрики в структуре словарных статей указывали и авторы концепции Словаря украинского языка в 20-ти

томах В. М. Русановский и В. А. Широков [11], что и нашло отражение в уже изданных первых трех томах указанной лексикографической работы.

Однако существующие определения, описанные признаки, критерии выделения, особенности функционирования, взаимоотношения с другими языковыми структурами зачастую, к сожалению, оказываются недостаточными при квалификации отдельных эквивалентов слова, которые иногда относят к классу фразеологических единиц. При отнесении эквивалентов слова (далее – ЭС) к фразеологическим единицам не учитываются существенные различия свойств рассматриваемых языковых элементов. Раздельно-оформленность фразеологизмов и ЭС действительно является их общей характеристикой, хотя истоки этого признака различны: генетической основой фразеологизма являются словосочетание [16; 1] и более сложные синтаксические образования [2; 5; 12; 9], ЭС же преимущественно происходят от форм слов с расчлененным строением типа *на рівних*, *без малого*, *для чого* и синтагматических соединений неполнозначных слов: *то...то*; *чи...чи*; *чи ба*. Поэтому и воспроизводимость ЭС и фразеологизмов различается: в первом случае воспроизводятся определенные формы слова, состоящие из нескольких графически отдельных компонентов, или синтагматически объединены неполнозначные слова, во втором – структуры, имеющие признаки внешней формы словосочетаний или более сложных чем словосочетания образований.

Акцентологические свойства ЭС и фразеологизмов тоже характеризуются определенными различиями. Фразеологизмы, как правило, имеют минимум два ударения, для ЭС два ударения скорее явление исключительное, они, как и слова, преимущественно характеризуются одним ударением. ЭС с тремя и более ударениями, как это имеет место для фразеологизмов, не фиксируются. Однако в целом признаком фразеологизма свойственна вариативность. Однако в целом фразеологизмам, наряду с устойчивостью структуры, свойственны и другие виды формального и лексического варьирования, собственно «диалектическое единство этих двух противоположных характеристик фразеологических единиц обеспечивает и функционирование, и постоянное развитие фразеологии» [8, с.10]. Вариативность ЭС чрезвычайно ограничена, предсказуема, поскольку главным образом связана с первичными формами, на базе которых создаются рассматриваемые единицы. ЭС, образованные от предложно-именных соединений, которые характеризовались вариативностью окончаний в определенных падежах, способны сохранять такой вид варьирования: *скільки раз (разів)*, *по полудню (полудні)* и т.д. Итак, если варьирование фразеологических единиц способствует «постоянному развитию фразеологии», то варьирование ЭС в основном свидетельствует о незаконченном характере их формирования и ещё живыми связями с образованиями, от которых они произошли. Создание ЭС происходит по определенным моделям. Последние, например, могут быть следующие: «предлог **в** + существительное в винительном падеже»: *в(у) лоб*, *в(у) струнку*, *у відкриті*, «предлог **на** + существительное в предложном падеже»: *на відшибі*, *на одчепі*, *на льоту*. Возможность создания ЭС по тем или иным моделям сближает их со словом (хотя модели, по которым происходит создание слов, отличаются от моделей ЭС).

На появление фразеологической единицы в языковой системе указывает дополнительное лексическое значение для определенных соединений по отношению к прямому значению его компонентов. По мнению А.В. Молоткова, «фразеологизм – это то или иное переосмысление конкретного словосочетания или предложения» [13, с. 9]. Переосмысление словосочетания связано с приобретением им признаков экспрессивности, именно «образность и

выразительность словосочетания является тем «магнитом», который притягивает к нему то дополнительное лексическое значение, превращающее словосочетание в фразеологизм [14, с. 67–68]. В ЭС же происходят только семантические сдвиги, определенные сдвиги значения, десемантизация полнозначных компонентов, а следовательно, и признаки экспрессивности их внутренней организации чужды.

Десемантизация полнозначных компонентов в составе ЕС, по нашему мнению, связана с относительно высокой частотностью их употребления. Очевидно, этот факт можно квалифицировать как еще одну причину, которая влияет на становление эквивалентов слова в языковой системе, поскольку именно наблюдения за непосредственным словоупотреблением «разрешает проблему с безошибочностью интуиции преимущественно вопреки традиционным решениям» [4, с. 183]. Фразеологические же единицы, как считают авторы «Проекта частотного словаря русских идиом», преимущественно малочастотные, кроме того, «многие фразеологизмы характерны лишь для некоторых типов дискурса и речевых жанров» [3, с. 30].

Подтверждению или отрицанию последних соображений будет способствовать обращение к национальным лингвистическим корпусам, в нашем случае – украинскому, в котором фиксируется более 54 млн словоупотреблений. В общем смысле лингвистический корпус – это представленный в электронном виде, большой по объему, унифицированный, структурированный, размеченный и филологически компетентный массив текстов на естественном языке, дополненный системой управления – в той или иной степени универсальными программными средствами для поиска и обработки различной лингвистической информации. В Украинском языково-информационном фонде НАН Украины в рамках работ по развитию Национальной словарной базы Украины создан Украинский национальный лингвистический корпус (далее УНЛК) [17]. Основные результаты теоретических исследований и обзор практических разработок в области лингвистических корпусов подробно представлены в коллективной монографии «Корпусная лингвистика» [15].

Основными направлениями УНЛК можно считать следующее:

- предоставление текстовой информации по определенным критериям;
- создание входных потоков лингвистической информации для разнообразных исследовательских систем;
- интеграция разноплановых лингвистически-программных средств обработки текстов в единой среде.

УНЛК разработан на базе созданной в Украинском языково-информационном фонде платформы ULISS (Ukrainian Lingua-Information Semantic Search). Особенностью данной платформы является возможность интеграции информационных и лингвистических сервисов в единой среде и поддержка режима виртуальных лабораторий. Сервис УНЛК аккумулирует функциональные возможности сервисов других информационно-лексикографических систем, а именно – электронной библиотеки, грамматического словаря, интегрированной лексикографической системы «Словари Украины», а также функций полнотекстового поиска. Ко всем перечисленным системам разработаны отдельные клиентские программы, которые обеспечивают интерфейсы работы с ними. УНЛК предоставляет единый интерфейс ко всем этим сервисам, создавая эффект работы с единой системой. Каждый указанный сервис отвечает за функциональность определенной лексикографической системы отдельно, а взаимодействие этих сервисов вместе в программном комплексе УНЛК создает необходимую среду обработки и хранения информационных ресурсов.

В системе УНЛК выделяют две основные функциональные подсистемы: библиографическую и полнотекстовую. Библиографическая часть представляет собой электронную библиотеку – коллекцию электронных ресурсов, которая является основой для разработки любого корпуса. Поэтому библиографическая подсистема служит инструментом для сбора, хранения, моделирования и использования естественной языковой информации в цифровом виде и составляет неотъемлемую часть УНЛК.

Вторая возможность доступа к объектам – это полнотекстовый поиск. Пользователь вводит поисковую фразу, задает максимально желаемое количество слов между поисковыми и выбирает дополнительные параметры полнотекстового поиска, а именно:

- учет порядка слов;
- поиск в определенном подмножестве объектов;
- использование процедуры лемматизации;
- использование синонимической лексикографической базы данных;
- использование определенных синонимических рядов;
- выбор грамматических параметров для каждого слова, входящего в поисковый фрагмент (см. рис. 1).

Результатом полнотекстового поиска является список библиографических описаний. Но в отличие от поиска по библиографии пользователь получает прямой доступ к каждой локализации поискового фрагмента в тексте, то есть ко всем контекстам, содержащим поисковый фрагмент. Выбрав источник, пользователь может просматривать контексты, где поисковый фрагмент выделен красным цветом. Размер (длину) контекста можно изменять. Каждое слово из контекста (не только поисковое) соединено со словарями: грамматическим, толковым и словарем синонимов. Для дальнейшей обработки все контексты или контексты определенного источника могут быть записаны в HTML-файл с указанием источника контекста, времени создания и поисковой фразы.

Система лингвистического корпуса является многоплановой и может иметь множество различных применений. В Украинском языково-информационном фонде она прежде всего используется как база источников лингвистической информации для создания фундаментальной академической многотомной лексикографической системы «Словарь украинского языка» (далее – СУМ), которая реализована как виртуальная лексикографическая лаборатория. Мы полагаем, что УНЛК может выполнять и в определенной степени функции эксперта при решении разнообразных лингвистических задач, в частности – разграничении эквивалентов слова и фразеологизмов.

Имеющиеся словари эквивалентов слова содержат около 1200 единиц, в СУМе фиксируются около 700 эквивалентов слова, которые регулярно пополняются и уточняются. Эти единицы и служат входом для текущих наборов, сформированных текстами указанных функциональных стилей, поскольку для установления статистической информации требуются данные общей частоты употребления ЭС, то есть их количество на миллион слов корпуса, а также активность указанных единиц в отдельных функциональных стилях.

Установлено, что ЭС характеризуются разной принадлежностью к частям речи, среди них выделяются:

- эквиваленты причастий: *до впаду, без перепочинку, з цікавістю,*
- местоимений: *один одного, де без кого, де з чим;*

- предлогов: *назустріч з; разом з; услід за;*
- союзов *тому що; не тільки, але й; тоді як;*
- ЭС, которые выполняют функцию частицы: *мало що; ось (он, от) як; ось (он, от) що;*
- эквиваленты междометий: *ой як; чи бач; от тобі й на;*
- ЭС, которые выполняют функции вставных слов: укр. *на щастя, до речі, може бути.*

Кроме того, к эквивалентам слова могут принадлежать элементы с несформировавшимися грамматическими признаками, в частности, так называемые предикативы типа *в ажурі, в моді* и т.п., а также цетеры (*і подібне, і таке інше*), обороты речи (*рідч у тому, що; факт той, що*). Все эти разряды ЭС, очевидно, характеризуются разной степенью частотности, что, конечно, требует экспертизы, которую можно осуществить с помощью Украинского лингвистического корпуса. Такого рода количественные подсчеты позволяют, с одной стороны, установить взаимоотношения относительно частотности употребления между фразеологизмами и ЭС, с другой – между различными лексико-грамматическими разрядами собственно эквивалентов слова.

Следующий этап, который бы способствовал разграничению (или, наоборот, объединению) фразеологизмов и ЭС, – это установление их частотности в произведениях отдельных авторов. В лингвистике давно уже замечено, что «использование фразеологии является индивидуальной особенностью идиолекта» [3, с. 30], однако, на наш взгляд, это не касается ЭС, хотя подтверждений последнему до сих пор нет. Создав в рамках Национального лингвистического корпуса корпусы текстов отдельных авторов, возможным представляется установить, функционируют ли ЭС как «фактор индивидуальный», что свойственно фразеологическим единицам, или ЭС – объективное явление общезыковой системы с конкретной формой выражения

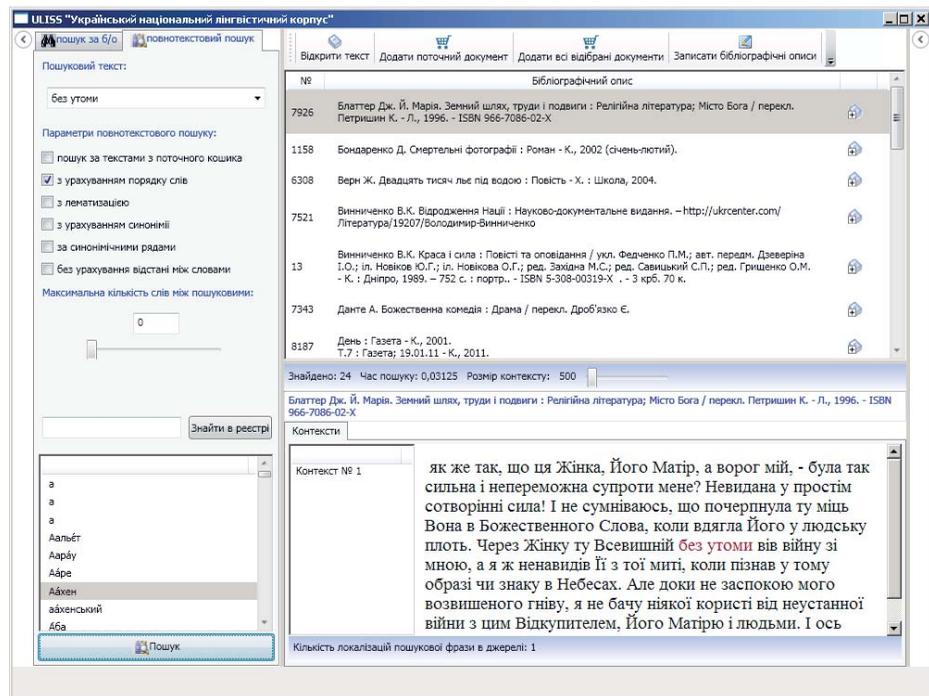


Рис. 1. Окно полнотекстового поиска в корпусе

и реальным содержанием. Возможно, здесь будут выявлены отдельные расхождения между отдельными разрядами ЭС, когда одни из них по этим параметрам будут тяготеть к собственно ЭС, а другие – к фразеологизмам.

Таким образом, результаты полнотекстового поиска эквивалентов слова в УНЛК будут способствовать решению как теоретических, так и прикладных задач. Установка частотности эквивалентов слова и фразеологизмов является новой теоретической и прикладной проблемой, над разрешением которой работают как лингвисты, так и разработчики информационных технологий.

ЛИТЕРАТУРА

1. *Алефіренко М. Ф.* Теоретичні питання фразеології. Харків: Вища школа. Вид-во при Харк. держ. ун-ті, 1987.
2. *Архангельский В. Л.* Устойчивые фразы в современном русском языке: Основы теории устойчивых фраз и проблемы общей фразеологии. Ростов / Дону, 1964.
3. *Баранов А. Н., Вознесенская М. М., Добровольский Д. О., Киселева К. Л., Козеренко А. Д.* Проект частотного словаря русских идиом // Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference «Dialog 2012». Bekasovo, 2012. P. 28–36.
4. *Кронгауз М. А.* Семантика. Москва: Ozon.ru, 2001.
5. *Кунин А. В.* Английская фразеология: Теоретический курс. М.: Высш. шк., 1970.
6. *Лучик А. А.* Словник еквівалентів слова української мови. Katowice: Wyd-wo US, 2006. 220 с.; *Лучик А. А.* Словник еквівалентів слова української мови. К.: Вид. дім «Киево-Могилянська академія», 2008.
7. *Лучик А. А.* Російсько-український і українсько-російський словник еквівалентів слова. К.: Довіра, 2003. 497 с.; *Лучик А. А., Антонова О. О., Дубровська І.* Українсько-польський словник еквівалентів слова. К.: Кассіопея, 2011. 312 с.; *Лучик А. А., Антонова О. О.* Польсько-український словник еквівалентів слова. К.: Кассіопея, 2013.
8. *Мокиенко В. М.* Современная фразеология. М.: Высшая школа, 1989.
9. *Молотков А. В.* Основы фразеологии русского языка. Л.: Наука, 1977.
10. *Рогожникова Р.П.* Словарь эквивалентов слова: наречные, служебные, модальные единства. Москва: Рус яз., 1991.
11. *Русанівський В. М., Широков В. А.* Інформаційно-лінгвістичні основи сучасної тлумачної лексикографії // Мовознавство. 2002. № 6. С. 7–48.
12. *Скрипник Л. Г.* Фразеологія української мови. К.: Наук. думка, 1973.
13. *Фразеологический словарь русского языка / Под ред. А.И. Молоткова.* 3 изд., стереотип. М.: Рус. язык, 1978.
14. *Чекман В. Н.* К вопросу о природе фразеологизмов // Вопросы фразеологии. Труды Самаркандского гос. ун-та. Новая серия. Вып. 219. Ч.1. Самарканд, 1972.
15. *Широков В. А.* та ін. Корпусна лінгвістика. К.: Довіра, 2005.
16. *Шмелёв Д. Н.* Современный русский язык. Лексика. М.: Просвещение, 1977.
17. <http://lcorp.ulif.org.ua/>.

Сведения об авторах:

Широков Владимир Анатольевич,

доктор технических наук, академик, директор Украинского языково-информационного фонда НАН Украины, основные научные интересы лежат в области формальных моделей языка и лингвистических технологий.

Лучик Алла Анатольевна,

доктор филологических наук, профессор кафедры общего и славянского языкознания национального университета «Киево-Могилянская академия», Киев, основные научные интересы лежат в области лексикографии, грамматики и лексикологии.

Остапова Ирина Вадимовна,

кандидат технических наук, научный сотрудник Украинского языково-информационного фонда НАН Украины, основные научные интересы лежат в области компьютерных лексикографических систем.

Яблочков Никита Миртович,

аспирант Украинского языково-информационного фонда НАН Украины, основные научные интересы лежат в области лингвистических технологий.