

ФИЛОЛОГИЯ

УДК 81'33

А.А. Баркович

КОРПУСНАЯ ЛИНГВИСТИКА: СПЕЦИФИКА СОВРЕМЕННЫХ МЕТАОПИСАНИЙ ЯЗЫКА

Статья посвящена исследованию и систематизации специфики современных метаописаний языка в контексте корпусной лингвистики. Рассмотрена проблематика формирования терминосистемы, интеграции корпусной лингвистики в лингвистическую парадигму, методологической релевантности корпусного подхода к широкому спектру исследований речи. Корпусная лингвистика обладает значительным потенциалом развития, реализация которого обусловлена металингвистическим обобщением и идентификацией референтной специфики.

Ключевые слова: корпусная лингвистика; корпусный дискурс; методология; парадигма; речевая практика; метаязык; металингвистика; метаописание.

Введение

Современные условия функционирования языка предполагают не только исследование образцов речи в отдельных хронологических или пространственных обстоятельствах, но и систематизацию речевой практики. Системный подход к изучению и обобщению речевой практики обуславливает видение языка как специфического единства. Традиционно считается, что только сплошной, без влияния субъективных факторов анализ эмпирического материала позволяет получить полную, структурированную картину объекта. Однако сегодняшний объект лингвистики в компьютерно-опосредованном пространстве и коммуникационно-опосредованном времени создается, сохраняется и реализуется в недоступном для изучения в «докомпьютерные» времена объеме. Конечно, задача тотального охвата речевой практики рассматривается сегодня – благодаря объективным критериям прикладной сферы – все более реалистично. При этом, благодаря новому «лингвоинформационному» формату, стали осязаемыми новые возможности изучения универсально-статистических и специально-предметных «выборок» языковых фактов. Появившись в 60-е гг. XX в., корпусная лингвистика существенно уточнила исследовательские приоритеты, создав принципиально новую методологическую основу для репрезентации речевой практики. Корпусная лингвистика как металингвистический инструмент – методология создания, использования и изучения корпусов для метаязыковой интерпретации закономерностей речевой практики посредством компьютерных инструментов – позволяет эффективно и успешно решать как «старые», так и «новые» проблемы изучения языка. Необходимость обобщения метаязыковой специфики корпусной лингвистики обусловлена ее высокой востребованностью, противоречивостью компетентной интерпретации, динамикой развития компьютерно-опосредованной коммуникации и другими факторами.

Интерпретация речевой практики постепенно стала синкретичной и междисциплинарной проблемной областью современной науки, а речь стала по-настоящему актуальным источником для качественно

новых научных обобщений. Вместе с тем значимость лингвистики как таковой не растворилась в контексте научных представлений о современной коммуникации. Собственно лингвистика всегда была важным резервом научной сферы: «...можно констатировать, что методы языкознания дают пример и даже становятся образцом для других дисциплин, что проблемы языка занимают сейчас самых разных специалистов, количество которых постоянно растет, и что общее направление мысли поощряет все гуманитарные науки работать в том же духе, что и лингвистика» [1. С. 21]. Высказанное Э. Бенвенистом в 70-х гг. XX в. наблюдение более чем актуально и сегодня. В «эпоху» компьютерно-опосредованной коммуникации становится все более очевидной зависимость от лингвистической интерпретации многих аспектов информационного и технического обеспечения социокультурного прогресса.

Лингвистический корпус в метаязыковом аспекте. В парадигме современных исследований языка корпусная лингвистика выделяется методологической универсальностью и эффективностью. Для корпусной парадигматики, изначально ориентированной на речевую практику, в современных условиях характерны четкая идентификация проблемной области, динамика и функциональность, аутентичность и высокая прецизионность результатов, богатство терминологического аппарата и востребованность обобщений как смежными дисциплинами, так и прикладной сферой в целом. Подобное уникальное для гуманитарной сферы сочетание характеристик обуславливает исключительную целесообразность корпусных возможностей в познавательной деятельности. Это не осталось незамеченным учеными, постоянно расширяющими сферу эффективного применения корпусной методологии.

De facto работа с корпусами – развитие квантитативно и эмпирически обоснованных обобщений – стала одним из основных векторов гуманитарных исследований, посредством корпусных исследований решаются действительно сложные задачи. Корпусная лингвистика полностью отвечает такому современному тренду «идеологии», по выражению В.А. Плунгяна, как формализация языка. Формализация языковых отношений,

репрезентация языковой системы в контексте компьютерного опосредования коммуникации, в том числе для создания искусственного интеллекта, стали мощнейшим стимулом максимально широкого метаязыкового охвата речевой практики.

Характер развития компьютерно-опосредованной коммуникации способствует быстрому накоплению текстовых массивов, зачастую дублирующих уже созданные «памятники» либо незначительно их модифицирующих. Сложно в данном контексте переоценить значимость эффективных инструментов исследования текстов. Корпусную лингвистику в сложившейся ситуации следует признать одним из наиболее существенных приобретений лингвистики за последние пятьдесят лет – уже не столько как подраздел или раздел языкознания, сколько в качестве широкой методологической базы интердисциплинарного характера: «...без обращения к методике корпусной лингвистики современная исследовательская практика редко обходится в принципе» [2. С. 45].

Определение корпуса как феномена – задача часто обсуждаемая, вклад в ее решение сделан Д. Синклером, Т. МакЭнери, Д. Байбером, А. Килгариффом, В.П. Захаровым, В.А. Плунгяном, В.Ш. Рубашкиным, А.Н. Барановым и другими учеными. Тем не менее по-прежнему важными в лингвистическом контексте представляются проблемы, касающиеся выявления понятийно-концептуальной специфики ключевой для всего «корпусного» лексико-семантического поля лексемы корпус и ее систематизированной метаязыковой рефлексии, метаязыковой кодификации развития леммы корпус и терминологического освоения накопленного потенциала метаописаний.

В понятийно-концептуальном аспекте необходимо отметить многозначность лексемы корпус, разные толковые словари фиксируют в среднем 6–8 ее основных значений, например:

1. м. Туловище человека или животного.

2. м. 1) Внутренняя опорная часть какого-л. здания, сооружения, на которую крепятся другие части; каркас.

2) а) Остов судна, самолета и т.п. с наружной обшивкой. б) Внешняя оболочка, покрытие какого-л. механизма, прибора и т.п.

3) Отдельное здание в ряду нескольких однотипных или обособленная часть большого здания.

3. м. Крупное войсковое соединение.

4. м. 1) Совокупность лиц какого-л. одного официального положения, одной специальности и т. п.

2) перен. Полное собрание, свод каких-л. текстов, основа словаря.

5. м. Закрытое среднее военно-учебное заведение (в Российском государстве до 1917 г.).

6. м. Название одного из размеров типографского шрифта (в полиграфии) [3].

Данный пример интересен как присутствием в перечне лексико-семантических вариантов корпуса «лингвистического» значения ‘полное собрание, свод каких-л. текстов, основа словаря’, так и его квалификацией как «переносного». Характеристика в словаре данного значения как переносного объясняется лишь

определенной новизной соответствующего денотативного основания. Вместе с тем не меньшей метафоричностью – относительно очевидного семантического дериванта ‘тело, туловище’ – отличаются и другие, более привычные значения корпуса: ‘остов’, ‘здание’, ‘учреждение’ и т.д. При этом нужно отдать должное автору словаря: данный словарь – один из немногих лексикографических источников, фиксирующих «лингвистическую» значимость корпуса, тем более в узком контексте (тип данного словаря – «толково-словообразовательный»).

На сайте интернет-ресурса «Словари и энциклопедии на Академике» [4], например, несколько десятков ссылок на словарные статьи о корпусе. Не все из них, правда, могут быть соотнесены с лингвистической спецификой. Тем не менее отдельные статьи о «лингвистическом» корпусе в русскоязычных словарях, как оказывается, нет, если не считать таковым упоминание о «Корпусе сочинений Дионисия Ареопажита» (в «Словаре книжников и книжности Древней Руси»). При этом принцип устройства данного «электронного» ресурса, несомненно, корпусный, не говоря уже о том, что составлением словарей собственно лингвисты и занимаются. Лексикографическая инновационность лексемы корпус – следствие неустойчивости и вариативности соотносимой с данным языковым знаком сигнификативной базы.

Данные общенародного языка, отраженные в толковых словарях, конечно, не дают полного представления о метаязыковой и, особенно, о металингвистической специфике современного языкознания. Тем не менее рассмотрение репрезентаций корпуса в толковых и специализированных словарях обеспечивает последовательную интерпретацию соответствующей значимости в контексте языковой системы, способствует верификации языковой единицы в среде речевой практики. Кроме того, особенности фиксации соответствующей семантики в разных словарях позволяют объективно судить о сформированной метаязыковой компетенции носителей языка как в рамках специализированного описания, так и в контексте узкого словоупотребления.

В функциональном аспекте характерно отсутствие в русском языке единой жесткой метаязыковой кодификации развития леммы корпус. До сих пор отмечается конкуренция двух деривационных парадигм лексемы корпус: используются как словоизменительные формы типа *ко́рпусы* и словообразовательные – типа *ко́рпусный*, так и альтернативные им формы типа *ко́рпуса*, *ко́рпусной* (для прилагательного мужского рода).

Речевая практика пока не склонила чашу весов в сторону какого-либо из вариантов деривационного развития лексемы «корпус». Как показывает корпусный дискурс, речевая деятельность, обусловленная практикой создания, использования и изучения корпусов – далеко не все предпрешено в конкуренции данных парадигматических вариантов:

«Бывало, идешь по цехам, людно: деревообделочный, механический, корпусной, литейный – везде шум и гром» (Борис Екимов. Котенок на крыше (2001));

«Корпусный цех перестал держать другие цеха, вышел из прорыва, из которого, по здравому рассуждению, выйти не мог в принципе» (И. Прусс. Эксперимент совсем не лабораторный // Знание – сила. 1988) [5] и т.д.

В приведенных примерах даже с одним и тем же определяемым словом цех использованы и прилагательное *корпусный* и его аналог – *корпусной*. Подобная ситуация типична для русского языка, в том числе относительно словосочетания атрибутивов *корпусный* (ж. р. *корпусная*) и *корпусной* (ж. р. *корпусная*) с лексемой лингвистика. В устном научном дискурсе встречаются как формы одного типа *корпусная лингвистика*, так и другого – *корпусная лингвистика*.

Об орфоэпической норме дериватов-прилагательных говорится в «Словаре правильной русской речи»: «При образовании прилагательных от имен существительных в них сохранялось то ударение, которое было свойственно производящему слову, т.е. существительному. Однако в XX в. наблюдается процесс перемещения ударения в прилагательных с корня на суффикс или, если ударение падает на первые слоги, – ближе к концу слова. Язык достаточно консервативен, чтобы быстро изменить произношение целого ряда слов, поэтому в речи определенное время сохраняется как традиционное произношение, так и новое. Этим и объясняется вариантность в произношении некоторых слов, например: *заводский* – *заводской*, *тормозный* – *тормозной*, *запасный* – *запасной* и т.п.» [6. С. 37]. В приведенный перечень примеров вполне можно было бы включить и диадру *корпусный* – *корпусной*. Но, разумеется, можно указать и на другие тенденции русскоязычного узуса: огромное количество заимствований, медленная их ассимиляция и, как следствие, тенденция к сохранению аутентичных экспонентов, что выражается, в том числе, в сохранении ударения как в самом заимствовании, так и в его дериватах. Лексема *корпус* – латинизм (*corpus* – ‘тело’), значения которого коррелируют во многих заимствовавших данную единицу языках (см., например, [7]). Актуальными и влиятельными в той или иной языковой среде оказываются как собственно языковые, так и внеязыковые факторы освоения заимствования: «...эксплицитно или имплицитно отраженные в речевом функционировании той или иной языковой единицы, диады, совокупности и т.д.» [8. С. 44].

С точки зрения номинационно-деривационной логики очевидна предпочтительность сохранения аутентичного ударения на первом слоге как для леммы существительного *корпус*, так и для соответствующих ей форм прилагательного – *корпусный*. Таким образом, формы типа *корпуса*, *корпусной* являются нормой, скорее, художественной литературы, постепенно приобретая оттенок архаичности.

При этом презентация самой соответствующей словарной статьи в вышеупомянутом словаре достаточно лаконична и, опять же, не содержит упоминаний о «лингвистической» специфике корпусного словоупотребления:

– *корпус*¹, ~са, мн. ~сы, ~сов, м. (туловище) *корпусный*;

– *корпус*² ~са, ~сов, м. (здание и др.) *корпусной* и *корпусный* [6].

Возможный ориентир в данном вопросе, способный закрепить самые экзотичные нормы, – словарная практика. Однако и словарные источники пока не закрепили единообразной нормативной репрезентации речевой практики. Например, «Русский орфографический словарь» в целом объективно отражает вариативность референтной речевой практики относительно самой леммы *корпус*:

– *корпус*¹, -а, мн. -ы, -ов (туловище; шрифт);

– *корпус*², -а, мн. -а, -ов (здание; воен.),

и соответствующих ей атрибутивов:

– *корпусной* (к *корпус*²);

– *корпусный* (к *корпус*¹) [9].

Однако к какому из указанных типов орфографических вариантов относится «лингвистический» *корпус* – прямых отсылок в данном словаре нет.

Классический лексикографический источник «Толковый словарь русского языка» С.И. Ожегова и Н.Ю. Шведовой содержит более развернутую экспликацию метаязыковой рефлексии леммы *корпус* (не упоминая при этом о «лингвистическом» *корпусе*):

КОРПУС, -а, мн. -а, -ов и -ы, -ов, м. 1. (мн. -ы, -ов). Туловище человека или животного. Крупный, тяжелый к. 2. (мн. -а, -ов). Остов, оболочка чего-н. К. корабля. К. часов. 3. (мн. -а, -ов). Отдельное здание в ряду нескольких или обособленная большая часть здания. Заводские корпуса. Боковой к. 4. (мн. -а, -ов). Крупное войсковое соединение из нескольких дивизий. Моторизованный, танковый, авиационный, воздушно-десантный к. 5. (мн. -ы, -ов). Совокупность лиц, объединенных общностью какого-н. официального положения. Дипломатический к. Депутатский к. Корреспондентский к. (в какой-н. стране: корреспонденты из разных стран). 6. (мн. -а, -ов). В дореволюционной России: среднее военно-учебное заведение. Кадетский к. Морской к. 7. Типографский шрифт с высотой литер около 3,76 мм. || прил. *корпусной*, -ая, -ое (к 4 и 6 знач.) и *корпусный*, -ая, -ое (к 1, 2, 3 и 7 знач.) [10].

Подобного стандарта в толковании – с небольшими вариациями – придерживаются и другие лексикографические источники русского языка (см., например, [3]). На словарные прецеденты русского языка равняются и словари украинского и белорусского языков (см., например, [11, 12]). Словарные статьи в приведенных словарях очень похожи по структуре на статьи из русскоязычных толковых словарей. Конечно, свои особенности в словарной интерпретации «*корпусной*» семантики присутствуют в отдельных лексикографических источниках. Так, например, в источнике «Великий тлумачний словник сучасної української мови» (укр. – Большой толковый словарь современного украинского языка) уже есть значение ‘повна збірка яких-небудь текстів’ (укр. – полное собрание каких-нибудь текстов) [11].

Если проанализировать вышеприведенные примеры, какая-либо рациональная металексическая мотивация дифференциации тех или иных вариантов леммы *корпус* и соответствующих дериватов практически отсутствует. Почему *корпус*–‘туловище’ и *корпус*–

‘здание’ дают отличные словоизменительные варианты по сравнению с корпусом-‘войсковым соединением’ и корпусом-‘учебным заведением’ – мотивированного ответа в русскоязычном (как и в украинском и белорусском) научном дискурсе нет.

Искусственность попыток уложить речевую вариативность корпуса (также, впрочем, как и огромного количества других языковых единиц) в «прокрустово ложе» грамматики хорошо видна на примере статьи из белорусского словаря «Граматычны слоўнік назоўніка» (бел. – Грамматический словарь существительного): «корпус м. НВ корпус, корпуса, корпусу, корпусам, корпусе; мн. НВ карпусы (3 ліч. 2, 3, 4 корпусы), карпусоў, карпусам, карпусамі, карпусах» [13]. В данном словаре предписывается с числительными 2, 3 и 4 использовать форму множественного числа корпусы (аналогично русскому варианту), а во всех других случаях – карпусы (русскоязычный эквивалент – корпусá). При этом какая-либо орфографическая дифференциация для форм единственного числа не предусмотрена: корпус, корпуса, корпусу, корпусам, корпусе. Сложно, конечно, представить, что среднестатистический носитель белорусского языка – как и русского или украинского – будет вспоминать при необходимости текст словарной статьи по поводу леммы корпус.

Справедливости ради нужно сказать, что и сам научный дискурс не отличается в терминологическом плане единообразием: часть лингвистов создает, использует и изучает *ко́рпусы*, а часть – корпусá, есть и те, кто спокойно относится к обоим вариантам леммы. Характерно, что словари лингвистических терминов достаточно противоречивы в определении «лингвистической» семантики корпуса. Рассмотрим два основных на сегодня источника лингвистической терминологии: «Словарь лингвистических терминов» О.С. Ахмановой и «Словарь лингвистических терминов» Т.В. Жеребило. Так, «Словарь лингвистических терминов» О.С. Ахмановой фиксирует «существенную» разницу между «примерной» и «все́й» совокупностями – «высказываний» или «произведений речи»:

КОРПУС (массив, текст) англ. corpus. 1. Примерная совокупность высказываний, отобранных для анализа и представленных в виде письменного текста, магнитофонной записи и т.п. 2. Вся сумма (совокупность) произведений речи, созданных коллективом носителей данного языка [14].

Другой источник, «Словарь лингвистических терминов» Т.В. Жеребило, содержит безупречное и лишь ассоциативно понятное определение корпуса:

КОРПУС ЯЗЫКА. Степень кодифицированности, наличия или отсутствия письменности и литературной нормы, включая нормы устной и письменной речи – орфоэпические, орфографические, грамматические, уровень междialeктной консолидированности, фактор литературной фольклорной и лингвокультурной традиции (В.П. Нерознак) [15].

Подобные метаязыковые описания, конечно, являются проблемой и нуждаются в коррекции с учетом современных представлений о корпусной методологии.

При этом современные компетентные суждения о корпусной деятельности остаются во многом дискуссионными, что существенно расширяет релевантную проблемную область. Так методологическая важность корпуса В.А. Плу́нгином ассоциируется не только с «идеологией», но и с возможностями генерировать самые широкие обобщения эмпирической практики: «Современные исследователи стали всё больше призывать лингвистику изучать факты, а не конструкты; рассуждать о свойствах наблюдаемых явлений, а не о свойствах моделей. Проблема же лингвистики в том, что язык, как известно, не наблюдаем и корпус как инструмент и как идеология представляет из себя, в некотором смысле, теоретическую абстракцию» [16. С. 8–9].

Не менее показательно мнение А.Д. Шмелёва, подчеркивающее важность непротиворечивой метаязыковой рефлексии, которая может быть распространена вплоть до языковой компетенции как таковой: «...использование корпуса может оказаться более всего плодотворным, когда исследователь критически оценивает данные, получаемые при помощи корпуса. В этом случае, обращаясь к собственной языковой компетенции, он имеет шанс обнаружить лакуны или неточности в существующих описаниях языка, в том числе тех, которые были использованы в системе поиска по корпусу. При таком подходе становится ясным, что ориентация на данные корпуса и на собственную языковую интуицию не противоречат друг другу, а скорее дополняют друг друга» [17. С. 264].

Актуальность рассмотренной проблематики является прямым следствием продолжающегося осмысления корпусной специфики на «философском» метаязыковом уровне, продуктивной лингвистической рефлексии. Металингвистическое освоение корпусной проблематики не является автономной и самодостаточной практикой, основываясь на широком круге метаязыковых обобщений.

Металингвистическая специфика корпусной парадигматики. Корпусная парадигматика обладает притягательностью как эффективная и высокодискретная область знаний, приумножая, в то же время, сомнения лингвистов, привыкших к классической «туманности» гуманитарных знаний. С другой стороны, высок интерес к корпусной методологии гуманитариев самой разнообразной компетенции и технических специалистов – инженеров и программистов. Речевая практика – универсальна, будучи отраженной и интерпретированной посредством корпусных инструментов, она является источником ценнейшей информации для широкого круга специалистов. Семантика корпуса в метаописаниях коррелирует не только с собственно лингвистической, но и с его общефилологической и даже общенаучной значимостью. Соответственно, рассмотрение методологической релевантности корпусной специфики обусловлено терминологической вариативностью корпуса: от общефилологической широкой значимости до прикладной узкоспециальной.

Современная концептуализация корпуса предполагает достаточно широкую репрезентацию референтной семантики. Типичным является следующее представление: «Корпус – это собрание текстов в

электронной форме, в котором можно осуществлять поиск слов, словосочетаний, грамматических форм, значений слов с помощью определенной поисковой системы» [5]. В широком смысле слова корпус – это практически любая совокупность языковых фактов, необязательно связанная с компьютерно-опосредованным представлением языка.

С узкоспециальным значением тоже особых проблем нет: лингвистический корпус (англ. text corpus) – снабженная разметкой и специальным программным обеспечением компьютерно-опосредованная совокупность текстов или текстовых фрагментов, предназначенная для статистически корректных исследований речи. С точки зрения информационных технологий корпус – это совокупность данных, единицами которой являются коммуникационно-релевантные паттерны семантики, обладающие иерархической макроструктурой. С точки зрения компьютерной науки корпус – это не просто коллекция отобранных по конкретной методике и представленных в электронном виде текстов из определенных сфер употребления языка (таких коллекций существует сейчас великое множество), но коллекция, категоризованная в аспекте интегральных характеристик каждого текста, и маркированная в плане формальных характеристик различных единиц его организации.

Разумеется, в контексте систематического исследования языка наиболее актуальными являются возможности именно лингвистического корпуса. В данной связи актуализируется целый ряд значимых научных проблем: определение пределов допустимого доминирования корпусных данных в сфере теоретического обустройства науки о языке; междисциплинарный консенсус самой корпусной деятельности – лишь частично собственно лингвистической или технической; необходимость формализации, репрезентации и интерпретации комплексной семантики языка и речи; учет и использование новой коммуникационной реальности, в частности Интернета, и др. В чем важность «инвентаризации» подобной проблематики? В специфике развития современной лингвистики, которую сегодня уже вполне можно дифференцировать на докомпьютерную и компьютерную (интуитивную и эмпирическую, информационную и семантическую и т.д.). Причем, если какое-то время компьютерная лингвистика представлялась зонтичной сферой для корпусной лингвистики (см., например, [18]), то сегодня уже нет сомнений в самостоятельности корпусной лингвистики в структуре языкознания (см., например, [19]). Более того, имеется достаточно оснований считать компьютерную парадигматику вспомогательной – для по-настоящему широко раскрывшейся корпусной методологии.

Универсальные возможности корпусной методологии обуславливают ее активное использование, особенно в комплексных исследованиях. Например, корпусный подход в дискурсивных исследованиях создал новый формат научной практики: «Это комплексное и тесное взаимодействие методов, подходов и инструментов запустило сближение двух областей [корпусной лингвистики и дискурс-анализа], корпусная лингвисти-

ка больше не находится на периферии анализа дискурса, а играет центральную роль» [20. С. 183].

До появления компьютеров, во времена доминирования структурной лингвистики, в условиях дефицита эмпирических данных, считалось вполне приемлемым делать далеко идущие выводы, основываясь на отдельных языковых фактах. Под такую методику даже подводилась соответствующая аргументация: «Не надо ждать накопления неограниченного множества высказываний, чтобы увидеть синтаксические правила, потому что они предшествуют появлению всего этого множества» [21. С. 16]. В современных условиях, при изобилии разного рода данных о языке, целесообразность подобного исследовательского подхода вызывает сомнения: почему бы не исследовать то новое, что стало доступным?

Нужно сказать, что без обращения к инструментарию корпусной лингвистики современная исследовательская практика обходится редко. В еще большей степени корпусно-зависимой является лингвистика компьютерно-опосредованной коммуникации. Корпусная методика многими исследователями признается важной и продуктивной: «Большие корпуса стали основным инструментом лингвистического исследования в последние два десятилетия. Этот период примерно соответствует времени закрепления речи в центре внимания лингвистических исследований» [22. С. 5]. С помощью корпусных технологий, с одной стороны, сегодня стали доступными для исследования и обобщения огромные массы текстов. С другой стороны, если учитывать в теоретических построениях все новейшие данные прикладной сферы, придется существенно корректировать теорию языка (см., например, [19]).

Не менее существенную проблемную ситуацию создает двойственная природа корпусной лингвистики: филологическая, корпуса содержат речевой материал, и техническая, корпуса работоспособны благодаря компьютерным технологиям. В результате корпуса создаются и совершенствуются специалистами разной компетенции. Синкретизм корпусной лингвистики требует формирования междисциплинарных команд, компетентных, в данном случае, в лингвистике, программировании и статистике: «...часто коллективы, специальные междисциплинарные коллективы, обладающие необходимым уровнем компетенции в лингвистике, программировании и статистике [выполняют корпусные исследования]...» [23. С. 53]. Возможно, нет для большинства филологов фобии страшнее, чем необходимость погрузиться в математику и информатику. Далеко не всегда приход технических специалистов в гуманитарную сферу сопровождается должным вниманием с их стороны к лингвистической сумме знаний. Проблема не нова, о ней писал еще Ф. де Соссюр: «...очевидно значение лингвистики для общей культуры: в жизни как отдельных людей, так и целого общества речевая деятельность является важнейшим из всех факторов. Поэтому немисливо, чтобы ее изучение оставалось в руках немногих специалистов», но, в то же время, владение языком ещё не означает профессиональной лингви-

стической компетенции: «...в действительности ею [лингвистикой] в большей или меньшей степени занимаются все; но этот всеобщий интерес к вопросам речевой деятельности влечет за собой парадоксальное следствие: нет другой области, где возникало бы больше нелепых идей, предрассудков, миражей и фикций» [24. С. 53]. Тем не менее в условиях закрепления междисциплинарного характера современной лингвистики данная проблема не стала менее актуальной – скорее, наоборот.

Недостатки традиционной лексико-центричной точки зрения на специфику системной презентации языка также были известны давно: «...слово, несмотря на все трудности, связанные с определением этого понятия, есть единица, неотступно представляющаяся нашему уму как нечто центральное в механизме языка; одной этой темы было бы достаточно для целого тома. Далее следовало бы перейти к классификации единиц низшего уровня, затем более крупных единиц и т.д. Таким образом, наша наука, определив элементы, которыми она оперирует, выполнила бы свою задачу полностью, сведя все входящие в ее ведение явления к их основному принципу. Нельзя сказать, что в лингвистике эта центральная проблема когда-нибудь уже ставилась и что все ее значение и трудность ее решения вполне осознаны; до сих пор в области языка всегда довольствовались операциями над единицами, как следует не определенными» [24. С. 143]. Безусловно, ограниченность презентации семантики лексическим уровнем стала еще более заметной в контексте уже сложившейся тенденции к формализации естественных языков и не менее активной практики по созданию искусственных языков – семиотических систем для сферы компьютерно-опосредованной коммуникации [25].

Опора на «привычный» языковой феномен, основанный на «иллюзиях» языковой картины мира, – слово – неудивительна в контексте повышенного интереса к перспективам формализации языковой системы. Однако общеизвестно, что, являясь, безусловно, лингвистической абстракцией, слово, точнее, лексическая единица, в метаязыковом контексте не является самостоятельным языковым артефактом, питаясь жизненной силой единого семантического поля языка, что справедливо отмечалось Ю.Д. Апресяном [26] и другими исследователями.

Корпусы в широком смысле слова, созданные в зависимости от целей исследования, существовали в лингвистике давно: начиная от словарей, собраний параллельных текстов (например, Библии) и заканчивая картотеками – личными и созданными научными учреждениями, как исходный материал для кейс-исследований [27]. При традиционной технологии сбора и обработки языковых данных агрегирование и обновление собранного материала являются нетривиальной задачей. Перманентная обработка картотеки, поиск нужных единиц и другие необходимые операции требуют много времени и усилий. Кроме того, «традиционная» лингвистическая техника совершенно не предусматривает доступ к языковым данным на расстоянии. И, конечно, новые «информационные» технологии и

технические средства (компьютерные системы, системы связи, системы мультимедиа) сделали языковые данные более доступными. Так, поскольку компьютерные технологии широко используются и в типографском деле, и в средствах массовой информации, агрегирование создаваемого в настоящее время «контента» существенно упростилось: большинство современных газет имеют версии, которые функционируют в компьютерно-опосредованном формате (все чаще определяемом как лингвоинформационный [28]). Быстрыми темпами нарастает зависимость речевой практики от коммуникационных сетей: сам Интернет является гиперкорпусом [19]. Современные технологии сканирования текста и эффективные программы расшифровки графической информации («изображения» текста) в собственно текстовый формат (текст как совокупность графем) позволили сделать практику перевода уже созданных «печатных» текстов в «электронный» формат массовой.

Практика верификации словарного состава языка в компьютерно-опосредованной среде коммуникационных сетей помогает синхронизировать работу лингвистов с требованиями времени. Подобная стратегия успешно реализована в серии словарей COBUILD, или Collins Birmingham University International Language Database (англ. – ‘Международная языковая база данных «Коллинз» и Бирмингемского университета’): «Все словари серии COBUILD базируются на информации, которую мы находим в Bank of English™ и Collins corpus» [29]. Если раньше словарная практика имела непререкаемый лингвистический авторитет, то сегодня – все чаще наоборот: словари базируются и выверяются по лингвистическим корпусам, средоточию образцов речевой практики. Характерно следующее наблюдение Н.Д. Голева: «В последнее время лингвисты, не довольные интуитивным решением проблемы формирования лексического состава словарей, все чаще обращаются к Национальному корпусу русского языка и данным Интернета, а также к лингвистическому эксперименту как более надежным источникам» [30. С. 185].

Современный корпусный формат становится все более востребованным, например, для создания программ машинного перевода: «В последнее время не может быть сомнений в важности или необходимости использования корпусов при переводе. Равным образом, онлайн-корпусы проявили себя как наиболее эффективный инструмент в контексте коротких сроков выполнения заданий и скорости, которые сейчас востребованы в индустрии переводов... Ad-hoc, специализированные корпусы, добытые из электронных ресурсов, доступных с помощью Интернета, доказали свой статус первоклассных источников текстов: как в качестве ценного инструмента в процессе принятия решений, так и при редактировании» [31].

Таким образом, речевая практика – как непосредственный предмет современной прикладной лингвистики – постоянно репродуцируется и реализуется в колоссальном разнообразии текстов. Практика создания ее метаописаний требует учета не произвольных выборок, а полного объема референтных данных. Само метаописание подразумевает репрезентацию объ-

екта в контексте и средствами вторичной семиотической либо абстрактной системы. Соответствующие корпусные обобщения востребованы, поскольку обладают объективным характером. При всей значимости для современной лингвистической парадигмы эмпирических данных развитие лингвистики как науки не может не сопровождаться систематизацией и обобщением первичных данных, их моделированием на уровне метаописаний, практикой репрезентации и интерпретации обобщений.

Выводы

Итак, сегодня на первый план постепенно выходят задачи не создания корпусной парадигматики: корпусная лингвистика создана и успешно функционирует. В современном научном контексте актуализируются задачи сохранения своеобразного методологического баланса: практики исследования языка и ее компетентной репрезентации, «метаязыковой» деятельности и «металингвистической» рефлексии. Метаязык и металингвистика, являясь близкими, но не эквивалентными понятиями, позволяют обоснованно дифференцировать многоаспектную лингвистическую специфику.

В то же время существенные сложности в инкорпорацию корпусных исследований в широкую метаязыковую практику возникают уже в процессе идентификации самого феномена лингвистического корпуса. Выявлена существенная специфика понятийно-концептуальной репрезентации ключевой для всего «корпусного» лексико-семантического поля лексемы корпус и ее систематизированной метаязыковой рефлексии, метаязыковой кодификации развития леммы корпус и терминологического освоения накопленного потенциала

метаописаний. В данном контексте показательны специфические особенности развития парадигмы лексемы корпус в русском, украинском и белорусском языках.

Потенциал корпусной методологии как основы создания объективных метаописаний сферы речевого функционирования обусловлен спецификой компетентной рефлексии корпусных исследований, являющейся дискуссионной и продуктивной. Металингвистическое освоение корпусной проблематики не является автономной и самодостаточной практикой, основываясь на широком круге метаязыковых обобщений. В данной связи рассмотрены металингвистическая интерпретация и верификация корпусной деятельности: определение пределов доминирования корпусных данных в сфере теоретического обустройства науки о языке; междисциплинарный консенсус корпусной деятельности; необходимость функциональной формализации, репрезентации и интерпретации семантики; адаптация корпусной парадигматики к условиям новой коммуникационной реальности, в частности Интернета, и др.

Корпусная методология – инструментарий, позволяющий комплексно исследовать языковую проблематику практически без изъятий; уникальная основа для решения сложных синкретичных проблем лингвистики, позволяющая избежать избирательной «системной» или «эмпирической» ориентации парадигматики. Метаязыковая значимость корпусной парадигматики подкреплена дискуссионностью рефлексии, прецизионностью и многоаспектностью полученных результатов, постоянным расширением сферы ее задействованности в научном дискурсе. Метаязыковая систематизация корпусной парадигматики – необходимая составляющая последовательной и аргументированной реализации ее потенциала.

ЛИТЕРАТУРА

1. Бенвенист Э. Общая лингвистика : пер. с фр. М., 1974. 447 с.
2. Баркович А.А. Методологический аспект изучения компьютерно-опосредованного дискурса // Вестник Нижегородского государственного лингвистического университета им. Н.А. Добролюбова. Н. Новгород, 2015. Вып. 30. С. 38–48.
3. Ефремова Т. Ф. Новый словарь русского языка: Толково-словообразовательный: Св. 136 000 слов. ст., ок. 250 000 семант. ед. : в 2 т. М., 2000. 2297 с.
4. Словари и энциклопедии на Академике. URL: <http://academic.ru> (дата обращения: 30.07.2015).
5. Национальный корпус русского языка. URL: <http://www.ruscorpora.ru> (дата обращения: 30.07.2015).
6. Соловьев Н.В. Словарь правильной русской речи: ок. 85 000 слов: более 400 коммент. М., 2008. 847 с.
7. Online Etymology Dictionary. Available at: <http://www.etymonline.com> (дата обращения: 30.07.2015).
8. Баркович А.А. Функциональность диады «коммуникационный – коммуникативный»: дискурсивный аспект // Вестник Томского государственного университета. Филология. 2015. № 5 (37). С. 37–52. DOI 10.17223/19986645/37/3.
9. Русский орфографический словарь / В.В. Лопатин ред. 2-е изд. М., 2007. 960 с.
10. Ожегов С.И., Шведова Н.Ю. Толковый словарь русского языка: 80 000 слов и фразеологических выражений. 4-е изд., доп. М., 1999. 944 с.
11. Бусел В.Т. Великий тлумачний словник сучасної української мови. Київ, 2005. 1728 с.
12. Тлумачальны слоўнік беларускай мовы: у 5 т. / К. Крапіва рэд. Мн., 1977–1984. 4087 с.
13. Граматычны слоўнік назоўніка / А.Я. Міхневіч, В.П. Русак, рэд. Мінск, 2008. 1379 с.
14. Ахманова О.С. Словарь лингвистических терминов. М., 1966. 608 с.
15. Жеребило Т.В. Словарь лингвистических терминов. 5-е изд., испр. и доп. Назрань, 2010. 486 с.
16. Плунгян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. 2008. № 16 (2). М. : Языки славянской культуры. С. 7–20.
17. Шмелёв А.Д. Языковые факты и корпусные данные // Русский язык в научном освещении. 2010. № 19 (1). М. : Языки славянской культуры. С. 236–265.
18. Баранов А.Н. Введение в прикладную лингвистику. М., 2001. 358 с.
19. Баркович А.А. Интернет-дискурс: компьютерно-опосредованная коммуникация. М., 2015. 288 с.
20. Flowerdew L. Corpus-based discourse analysis // The Routledge handbook of discourse analysis / J.P. Gee and M. Handford eds. London: Routledge, 2012. P. 174–188.
21. Леви-Стросс К. Мифологии : в 4 т. [пер. с фр.]. СПб., 1999. Т. 1. Сырое и приготовленное. 406 с.
22. Mauranen A. Speaking professionally in an L2: Issues of corpus methodology // Variation and Change in Spoken and Written Discourse: Perspectives from corpus linguistics / J. Bamford, S. Cavalieri, G. Diani eds. Amsterdam: John Benjamins Publishing Company, 2013. Vol. 21 of Dialogue Studies. P. 5–32.

23. McEnery T., Hardie A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge, 2012. 294 p.
24. Соссюр Ф. де. Курс общей лингвистики // Труды по языкознанию. М.: Прогресс, 1977. С. 31–273.
25. Баркович А.А. Компьютерно-опосредованная коммуникация: потенциал металексической значимости // Ученые записки Петрозаводского государственного университета. 2015. № 7 (152). Общественные и гуманитарные науки. Петрозаводск: ФГБОУ ВПО «Петрозаводский государственный университет (ПетрГУ)», 2015. С. 38–43.
26. Апресян Ю.Д. Избранные труды: в 2 т. М.: Языки русской культуры. 1995. Т. I: Лексическая семантика. Синонимические средства языка. 464 с.
27. Баркович А.А. Лексічны патэнцыял беларускай мовы ў святле корпуснай лінгвістыкі. Мінск, 2012. 158 с.
28. Баркович А.А. Лингвоинформационная специфика компьютерно-опосредованной коммуникации: структурный аспект // Вестник Волгоградского государственного университета. Сер. 2. Языкознание. 2015. № 2 (26). С. 114–120. DOI: <http://dx.doi.org/10.15688/jvolsu2.2015.2.16>
29. About the Collins Corpus and the Bank of English™, available at: <http://www.mycobuild.com/about-collins-corpus.aspx> (дата обращения: 30.07.2015).
30. Голев Н.Д. Лексическая реализация как функциональная характеристика словообразовательной системы русского языка и количественные параметры ее описания // Осьмнадцать: сб. науч. ст. к 80-летию И.С. Улуханова; отв. ред. М.А. Малыгина. М., 2015. С. 183–193.
31. Corpus Pastor G., Seghiri M. Specialized Corpora for Translators: A Quantitative Method to Determine Representativeness // Translation Journal. 2007. Vol. 11, Num. 3. Available at: <http://www.bokorlang.com/journal/41corpus.htm> (дата обращения: 30.07.2015).

Статья представлена научной редакцией «Филология» 22 марта 2016 г.

CORPUS LINGUISTICS: THE SPECIFICS OF MODERN META-DESCRIPTIONS OF LANGUAGE

Vestnik Tomskogo gosudarstvennogo universiteta – Tomsk State University Journal, 2016, 406, 5–13.

DOI: 10.17223/15617793/406/1

Barkovich Aliaksandr A. Belarusian State University (Minsk, Republic of Belarus). E-mail: albark@tut.by

Keywords: Corpus Linguistics; corpus discourse; methodology; paradigm; speech practice; meta-language; metalinguistics; meta description.

The article is devoted to research and systematization of the specifics of modern meta descriptions of language in the context of Corpus Linguistics. *Corpus Linguistics* as a metalinguistic tool is the methodology of creation, use and study of corpora for meta-language interpretation of the speech practice regularities by computer means. The need to summarize meta-language specifics of Corpus Linguistics is due to its being in high demand, contradictory competent interpretation, dynamics of computer-mediated communication and other factors. The issues of term system formation, Corpus Linguistics integration into the linguistic paradigm, methodological relevance of the corpus approaches to a wide range of speech practice research works were described. The essential specifics of notional and conceptual representation of the key lexeme *corpus* for the whole “corpus” lexical-semantic field, its *meta-language* reflection and codification of the lemma *corpus* development and terminological absorption of the accumulated metadescriptions potential were revealed. *Metalinguistic* interpretation and verification of corpus activities were considered: the defining of the corpus data dominance limits in theoretical arrangement of the language science; interdisciplinary consensus of the corpus activities; the need for functional formalization, representation and interpretation of the semantics; the adaptation of corpus paradigmatics to the conditions of new communication reality, in particular the Internet, and others. Speech practice as the direct object of modern Applied Linguistics is constantly reproduced and realized in an enormous variety of texts. The practice of creation of its meta descriptions requires taking into account not only random samples, but the full scope of the reference data. *Meta description* by itself implies the object representation in the context and by means of a secondary semiotic or abstract system. Interpretation of speech practice gradually became a syncretic and interdisciplinary problem domain of modern science, and the speech has become a truly relevant source for qualitatively new scientific generalizations. In this context, the concept of *corpus discourse* is relevant as speech activities due to the practice of creating, using and studying the corpora. Thus, corpus methodology is the instrumentarium that allows exploring the language problems in a complex mode almost without exception; a unique basis for solving complex syncretic linguistic problems that allows avoiding the selectivity of “system” or “empirical” paradigmatic orientation. Meta-language systematization of the corpus paradigmatics is a necessary component of its consequential and reasoned potential realization.

REFERENCES

1. Benveniste, E. (1974) *Obshchaya lingvistika* [General linguistics]. Translated from French. Moscow: Progress.
2. Barkovich, A.A. (2015) Metodologicheskii aspekt izucheniya komp'yuterno-oposredovannogo diskursa. [The Study of Computer-Mediated Discourse: The Methodological Aspect]. *Vestnik Nizhegorodskogo gosudarstvennogo lingvisticheskogo universiteta im. N.A. Dobrolyubova*. 30. pp. 38–48. (In Russian).
3. Efremova, T.F. (2000) *Novyy slovar' russkogo yazyka: Tolkovo-slovoobrazovatel'nyy: Sv. 136 000 slov. st., ok. 250 000 semant. ed.: v 2 t.* [New Dictionary of the Russian language: Explanatory-derivational: 136,000 entries. c. 250000 semantic units. In 2 v.]. Moscow: Russkiy yazyk.
4. Dictionaries and encyclopedias at Academic.ru. [Online]. Available from: <http://academic.ru>. (Accessed: 30th July 2015).
5. The Russian National Corpus. [Online]. Available from: <http://www.ruscorpora.ru>. (Accessed: 29th May 2015). (In Russian).
6. Solov'ev, N.V. (2008) *Slovar' pravil'noy russkoy rechi: ok. 85 000 slov: bolee 400 komment* [Dictionary of correct Russian speech: about 85 000 words: more than 400 commentaries]. Moscow: AST: Astrel.
7. Online Etymology Dictionary. [Online]. Available from: <http://www.etymonline.com>. (Accessed: 30th July 2015).
8. Barkovich, A.A. (2015) Functionality of the “communicational-communicative” dyad: discursive aspect. *Vestnik Tomskogo gosudarstvennogo universiteta. Filologiya – Tomsk State University Journal of Philology*. 5 (37). pp. 37–52. (In Russian). DOI 10.17223/19986645/37/3
9. Lopatin, V.V. (ed.) (2007) *Russkiy orfograficheskii slovar'* [Russian spelling dictionary]. 2nd ed. Moscow: Institut russkogo yazyka imeni V.V. Vinogradova.
10. Ozhegov, S.I. & Shvedova, N.Yu. (1999) *Tolkovyy slovar' russkogo yazyka: 80 000 slov i frazeologicheskikh vyrazheniy* [Dictionary of the Russian language: 80 000 words and idiomatic expressions]. 4th ed. Moscow: Azbukovnik.
11. Busel, V.T. (2005) *Velikiy tumachnyi slovník suchasnoï ukrains'koï movi* [Great Dictionary of Modern Ukrainian language]. Kiïv: Perun. (In Ukrainian).
12. Krapiva, K. (ed.) (1977–1984) *Tumachal'ny sloŭnik belaruskay movy: u 5 t.* [Explanatory Dictionary of Belarusian language: in 5 volumes]. Minsk: Vydavectva “Belaruskaya saveckaya yencyklapedyja” imya Pyatrusya Broŭki. (In Belorussian).

13. Mihnevich, A.Ya. & Rusak, V.P. (eds) (2008) *Gramatychny sloŭnik nazoŭnika* [Grammar Dictionary of Noun]. Minsk: Belaruskaya navuka. (In Belorussian).
14. Akhmanova, O.S. (1966) *Slovar' lingvističeskikh terminov* [A Dictionary of Linguistic Terms]. Moscow: Sovetskaya entsiklopediya.
15. Zherebilo, T.V. (2010) *Slovar' lingvističeskikh terminov* [Dictionary of linguistic terms]. 5th ed. Nazran: Pilgrim.
16. Plungyan, V.A. (2008) Korpus kak instrument i kak ideologiya: o nekotoryh urokakh sovremennoy korpusnoy lingvistiki [The corpus as an instrument and as an ideology: some of the lessons of modern Corpus Linguistics]. *Russkij jazyk v nauchnom osvешhenii*. 16 (2). pp. 7–20.
17. Shmelyov, A.D. (2010) *Yazykovye fakty i korpusnye dannye* [Language facts and corpus data]. *Russkij jazyk v nauchnom osvешhenii*. 19 (1). pp. 236–265.
18. Baranov, A.N. (2001) *Vvedenie v prikladnuyu lingvistiku* [Introduction to Applied Linguistics]. Moscow: Editorial URSS.
19. Barkovich, A.A. (2015) *Internet–diskurs: komp'yuterno-oposredovannaya kommunikatsiya* [Internet discourse: computer-mediated communication]. Moscow: Flinta: Nauka.
20. Flowerdew, L. (2012) Corpus-based discourse analysis. In: Gee, J.P. & Handford, M. (eds) *The Routledge handbook of discourse analysis*. London: Routledge.
21. Levi-Strauss, C. (1999) *Mifologiki. V 4-h tt.* [Mythologies. In 4 volumes]. Translated from French. Vol. 1. St. Petersburg: Universitetskaya kniga.
22. Mauranen, A. (2013) Speaking professionally in an L2: Issues of corpus methodology. In: Bamford, J., Cavalieri, S. & Diani, G. (eds) *Variation and Change in Spoken and Written Discourse: Perspectives from corpus linguistics*. Vol. 21 of Dialogue Studies. Amsterdam: John Benjamins Publishing Company.
23. McEnery, T. & Hardie, A. (2012) *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press.
24. Saussure, F. de. (1977) Kurs obshchey lingvistiki [Course in General Linguistics]. In: Saussure, F. de. *Trudy po yazykoznaniiyu* [Works on linguistics]. Translated from French. Moscow: Progress.
25. Barkovich, A.A. (2015) Computer-mediated communication: potential of metalexical significance. *Uchenye zapiski Petrozavodskogo gosudarstvennogo universiteta. Obshhestvennye i gumanitarnye nauki – Proceedings of Petrozavodsk State University. Social Sciences and Humanities*. 7 (152). pp. 38–43. (In Russian).
26. Apresyan, Yu.D. (1995) *Izbrannye trudy: v 2 t.* [Selected Works: in 2 volumes]. Vol. 1. Moscow: Yazyki russkoy kul'tury.
27. Barkovich, A.A. (2012) *Leksichny potentsyial belaruskay movy ŷ svyatile korpusnay lingvistyki* [Lexical potential of the Belorussian language in the light of corpus linguistics]. Minsk: RIVSH. (In Belorussian).
28. Barkovich, A.A. (2015) Linguistic and informational specifics of computer-mediated communication: structural aspect. *Vestnik Volgogradskogo gosudarstvennogo universiteta. Seriya 2. Jazykoznanije – Science Journal of Volgograd State University. Linguistics*. 2 (26). pp. 114–120. DOI: <http://dx.doi.org/10.15688/jvolsu2.2015.2.16>. (In Russian).
29. About the Collins Corpus and the Bank of English™. [Online]. Available from: <http://www.mycobuild.com/about-collins-corpus.aspx>. (Accessed: 30th July 2015).
30. Golev, N.D. (2015) Leksicheskaya realizatsiya kak funkcional'naja harakteristika slovoobrazovatel'noj sistemy russkogo jazyka i kolichestvennye parametry ee opisaniya [Lexical implementation of the functional characteristics of word-formation system of the Russian language and quantitative parameters of its description]. In: Malygina, M.A. (ed.) *Osm' desyat': Sbornik nauchnykh statey k 80-letiyu I.S. Ulukhanova* [Eighty: A collection of articles on the 80th anniversary of I.S. Ulukhanov]. Moscow: Azbukovnik.
31. Corpas Pastor, G. & Seghiri, M. (2007) Specialized Corpora for Translators: A Quantitative Method to Determine Representativeness. *Translation Journal*. 11 (3). Available from: <http://www.bokorlang.com/journal/41corpus.htm>. (Accessed: 30th July 2015).

Received: 22 March 2016