

# СИНТЕЗ КОНЦЕПЦИЙ В ЛИНГВИСТИКЕ: ДИСКУРСИВНЫЕ И КОРПУСНЫЕ ИССЛЕДОВАНИЯ



**Boris Yu. Norman**<sup>1</sup>

Belarusian State University, Minsk

**Mikhail Yu. Mukhin**<sup>2</sup>

Ural Federal University, Ekaterinburg

## INTERDISCIPLINARY TRENDS OF MODERN LINGUISTICS

The active development of interdisciplinary research methods in linguistics began in the mid 20<sup>th</sup> century. Common interests with sociology, psychology, philosophy and other disciplines brought about the emergence of new conceptions and research areas. Anyway, linguistics, the same as philology in general, is known to originate from philosophy as a more general humanities discipline. In recent decades, efficient interaction between linguistics and informatics has become a salient interdisciplinary phenomenon. On the one hand, scholars have received programs for sound processing, database management systems, and tools for designing automatic speech analysis and synthesis. On the other hand, the findings of theoretical linguistics have turned out to be utterly urgent for communication modeling in information systems, and such vanguard research areas as machine translation, information search, and computer assisted learning have determined the narrow interpretation of interdisciplinarity in linguistics as sharing a common field of research with information technologies and not any other area of knowledge.

Corpus linguistics is a most important “driver” of the modern linguistic thought. This academic school is over sixty years old, but its influence upon socio-humanitarian research became really evident in the 1990s. Today, the development of not only linguistics but also various academic and applied spheres, interested in large text databases, is impossible without involving corpora. On the background of development of “classical” corpus databases one can witness the creation of new semantically-commented, multimedia parallel corpora and projects with hierarchical mark-up on the level of the phrase, discourse, etc., which can be considered new trends of corpus linguistics. The analytical potential of modern corpora is determined by linguistic, chronological and functional-textual mark-up parameters, and even traditional philological studies have usually included elements of “statistical corpus graphics”: schemes, graphs, diagrams, tables, etc.

The sophistication of corpus-based methods is closely associated with the new spiral of development of stylometry – a discipline presup-

<sup>1</sup> Boris Justinovich Norman, Guest Editor, Doctor of Philology, Professor of Department of Theoretical and Slavic Linguistics, Belarusian State University.

<sup>2</sup> Mikhail Yurievich Mukhin, Guest Editor, Doctor of Philology, Professor of Department of Fundamental and Applied Linguistics and Textology, Ural Federal University.

posing performance of computer assisted text modeling. With reference to the study of works of fiction, stylometry can be understood as a kind of new philology. Without rejecting traditional philological achievements, stylometry works out methods that make it possible to explore multiple textual phenomena, which can be overlooked in usual (even professional) reading of a work of fiction. This is true about the lexical and grammatical composition of the texts and the combinability of its elements – up to letters and other graphic signs.

New vectors can be seen in the further development of socio- and psycholinguistic research. Today, language, text and corpus materials provide multiple opportunities to investigate psychological issues of the person, the dynamics of developmental change and to monitor cognitive processes. Application of experimental methods, which allow modeling the conditions of emergence of a language fact attracting scholarly attention, is an undisputable advantage of psycholinguistic studies. Language and social conflict management, exploration of speech mechanisms in the development of various diseases, developmental specificity of cognitive structures, use of eye tracking technologies (measuring eye movement) and virtual reality to study text perception, psycho- and sociolinguistic characteristics of social networking site users are only some instances of the issues solved at the intersection of socio-humanitarian and natural disciplines. Social media provide constant sources of big data and this peculiarity has become more salient since the onset of the pandemic – the moment of significant qualitative and quantitative change in electronic communication.

The present-day research interest in the problems of connection between language and culture (material and spiritual) is not arbitrary: it has manifested itself more vividly in the ideas of cognitive linguistics, linguoculturology, studies of “conceptosphere” and “linguistic worldview”, etc. In all these cases, we come across boundary intermediate phenomena, which enrich our total knowledge about man.

Thus, interdisciplinary linguistics can explore synthetic material (both textual and multimedia), correlate linguistic and art graphics, and study interaction between various semiotic systems and transformation of information while transferring

from one system into another. Such is, for instance, the specificity of text transformation under various discourse conditions (fiction prose – scenario – screenplay – transcript), the problem of reflection of spontaneous oral speech in a written (including fiction) text, specificity of interaction of speech and gestures, etc. In other words, modern linguistics sets the tasks to explore sophisticated complex objects, the study of which needs additional methodological resources.

On the background of such opportunities, researchers are also faced with the basic problem of unchangeability and stability of the linguistic apparatus in the epoch of interdisciplinary methods. Isolation and heterogeneity of scholarly-practical vectors and the abundance of interdisciplinary ties raise the question of returning to a holistic and self-sufficient methodology and opposing blurring linguistic objects. Unfortunately, representatives of related fields, who do not have basic philological education, often carelessly use modern linguistic resources and, as a result, make wrong conclusions, which may compromise linguistic methods in the eyes of the non-professional community.

Nevertheless, looking for new research paradigms and keeping in touch with related spheres, linguistics has not lost its object of investigation. Both classical sources (text, dictionary, grammar) and modern resources (corpus, electronic thesaurus, search system) contribute to the situation when linguistic research becomes more objective and precise in the description of various forms of communication.

This section of the journal *Philological Class* presents interdisciplinary opportunities of the modern theory of language, which are realized on the borderline between linguistics and other humanitarian and natural scientific spheres of knowledge.

The article by B. Yu. Norman (Belarusian State University, Minsk) proceeds from the assumption about the presence in the native speaker's mind of a cognition about a syntactical “optimum” that allows using the linguistic units to maximum advantage to achieve a certain communicative effect: it is a unity of intuitive (“naïve”) knowledge about the parameters of the sentence, organization of its structure, rules of correspondence of lexical and grammatical meanings, etc. Deviations from this optimum,

manifested in a higher frequency of speech realization of a certain phenomenon, can become separate objects of not only scholarly research but of fiction text parody as well. The linguistic problems touched upon (word order, unmotivated ellipsis, parcellation, etc.) are interwoven with those of literary studies (parody as a genre) and psychological analysis (saliency, i.e. identifiability including cases in which it is not based on opposition).

The article by E. Yu. Protasova (The University of Helsinki, Finland) deals with analysis of polycode structures found in advertisements and signs displayed in various regions of Russia. On the one hand, the modern linguistic landscape should demonstrate features of multiculturalism – especially in such a multinational country as Russia. On the other hand, statistically reliable research material (several thousand photos) shows that the signs collected in various regions do not reflect the real language skills of their regional native speakers. The native language skills suffer from the influence of the state language of the Russian Federation and the specific impact of the English-language (and in some regions Chinese-language) culture. At the same time, the material reflects the life priorities of common people and their everyday and cultural needs. Specifically, the author draws a conclusion that the creation of polylinguistic (hybrid) graphic constructs can be considered a tool and index of internationalism and cultural diversity, especially if a large number of various ethnic groups live side by side in one and the same country.

The study of media text by E. N. Basovskaya (Russian State University for the Humanities, Moscow) and T. A. Vorontsova (Chelyabinsk State University, Chelyabinsk) looks at the correlation of phatic and informational components in professional and user-written movie reviews and comments. The scholars are interested to know if it is possible to reduce the “information density” of the utterance via enhancing the means of establishing and keeping contact with the reader-interlocutor or how to imitate real communication by means of a media text. Their analysis shows that the authors of reviews and comments solve the problem of keeping in touch with virtual addressees by various means – for example, precedent phenomena, different methods of self-presentation, vi-

vid stylistic devices (irony, etc.) – and thus effect communicative tuning in to “their” interlocutor. The authors arrive at a conclusion that the enhancement of the phatic component in the text leads to a certain information diversification and activation of communication with “their own reader”. Such a communication is similar to a conversation on a train, when the process of communication itself turns out to be more important than the information goals.

A. V. Kolmogorova (Siberian Federal University, Krasnoyarsk) carries out a study targeted at formalized identification (“detection”) of emotional dominants in the texts of social networking sites users. Emotive semantics modeling allows organizing computer assisted learning with the view of automation of recognition of the emotive constituent of the text. The article considers the issues of the difference in degree of formalization regarding various emotions (sadness, joy, anger, surprise, shame, etc.), which influences the effectiveness of computer assisted learning. And specifically, juxtaposed are texts in which we can detect manifestations of sadness and anger: sadness can hardly be recognized automatically, whereas anger, on the contrary, is detected rather well. The study employs crowdsourcing to create the initial database of emotive features; control tests were offered to two thousand informants, whose opinions formed the basis for further computer assisted learning. The author formulates social scenarios of prototypical experiencing emotions and actually suggests new technologies of lexico-semantic analysis.

The article by M. V. Kopotev (The University of Helsinki, Finland; Higher School of Economics, Saint Petersburg) defines the role of corpus linguistics in the solution of various problems of modern philology. The author characterizes various kinds of corpus analysis: *corpus-informed analysis* actually performs the illustrating research function; *corpus-based analysis* presupposes interpretation of statistical data; *corpus-driven analysis* leads to the classification of linguistic phenomena that seems to be performed without the researcher – an interpreter of large data arrays. The author makes conclusions about the contribution of corpus linguistics to the development of the general theory of language emphasizing the conventional nature of the modern dichotomy of language and speech. Investigation of speech

as a special kind of activity is of utmost importance for linguistics. In the first place, it does not focus on modeling the rules of interaction of linguistic units, but concentrates on the parameters of their usage “constituting an uninterrupted continuum”.

The study carried out by M. Yu. Mukhin and N. Yu. Mukhin (Ural Federal University named after the First President of Russia B. N. Yeltsin, Ekaterinburg) implements corpus stylometric methodology with reference to lexical combinability in the Russian prose of the 19<sup>th</sup> century. The article suggests a model of comparison of the peculiarities of authored syntagmatics, consisting of specific units – lexical bigrams – comprising two words used in a common phrasal context. The aim of the study is to systematize original authored word combinations typical of the creative activity of L. N. Tolstoy, F. M. Dostoevskiy, A. P. Chekhov, I. S. Turgenev and I. A. Goncharov. Such examples as “*oblokotit’ golovu*” (L. N. Tolstoy),

“*blagodarnye glaza*” (A. P. Chekhov) or “*nervicheskiy chelovek*” (I. S. Turgenev) are singled out with the help of lexico-statistical comparison of bigrams incorporating a word often used by all authors (*golova, glaza, chelovek*, etc.). The authors present their syntagmatic model as a systematized set of such word combinations which are found in the works of fiction of one writer and are not typical of other writers. The article combines findings in the field of linguistics, literary studies, statistics and personology (personal psychology). In terms of further research, the material can be presented as a unique writers’ syntagmatic dictionary.

The common aspects of the articles included in this section show that the ideas of building bridges between different scientific paradigms “are in the air”. We believe that the main interdisciplinary trends of modern linguistics consist in the study of synthetic objects, extralinguistic dimensions and the interpretation of big data in order to solve traditional and novel problems.

**Борис Юстинович Норман<sup>1</sup>**

Белорусский государственный университет, Минск

**Михаил Юрьевич Мухин<sup>2</sup>**

Уральский федеральный университет, Екатеринбург

## МЕЖДИСЦИПЛИНАРНЫЕ ТРЕНДЫ СОВРЕМЕННОЙ ЛИНГВИСТИКИ

Развитие междисциплинарных методов в лингвистике активно началось в середине прошлого века. Общие интересы с социологией, психологией, философией и другими науками стали причиной появления новых концепций и научных направлений. Впрочем, лингвистика и в целом филология, как известно, зародились в свое время в недрах философии как наиболее общей гуманитарной науки. В последние десятилетия очевидным междисциплинарным феноменом стало эффективное взаимодействие лингвистики и информатики. С одной стороны, гуманита-

рии получили программы обработки звука, системы управления базами данных, средства разработки автоматического анализа и синтеза речи. С другой стороны, данные теоретической лингвистики оказались крайне востребованными при моделировании коммуникации в информационных системах, и такие флагманские направления, как машинный перевод, информационный поиск, машинное обучение, определили узкое понимание междисциплинарности в лингвистике как пересечения именно с информационными технологиями.

<sup>1</sup> Борис Юстинович Норман, приглашенный редактор, доктор филологических наук, профессор кафедры теоретического и славянского языкознания, Белорусский государственный университет.

<sup>2</sup> Михаил Юрьевич Мухин, приглашенный редактор, доктор филологических наук, профессор кафедры фундаментальной и прикладной лингвистики и текстоведения, Уральский федеральный университет.

Важнейшим «драйвером» современной лингвистической мысли является корпусная лингвистика. Этому направлению уже более шестидесяти лет, однако его влияние на социогуманитарные исследования в полной мере проявилось начиная с 90-х годов XX в. Сегодня без привлечения корпусов немыслимо развитие не только языкознания, но и самых разных академических и прикладных сфер, проявляющих интерес к большим текстовым данным. На фоне совершенствования уже «классических» корпусных баз трендами развития корпусной лингвистики можно считать создание новых семантически аннотированных, мультимедийных, параллельных корпусов, проектов с иерархической разметкой на уровне фразы и дискурса и т. д. Аналитические возможности современных корпусов определяются лингвистическими, хронологическими и функционально-текстовыми параметрами разметки, и даже в филологические исследования, которые осуществляются в традиционном ключе, принято включать элементы «статистической корпусной графики»: схемы, графики, диаграммы, таблицы и т. п.

С развитием корпусных методов тесно связан новый виток развития стилометрии (или стилеметрии) – направления, предполагающего проведение компьютерного моделирования текста. В применении к изучению литературных произведений стилометрию в определенном смысле можно понимать как новую филологию. Не отвергая традиционные филологические достижения, стилометрия разрабатывает методы, позволяющие исследовать множественные текстовые явления, которые не всегда заметны при обычном (даже профессиональном) прочтении художественного произведения. Это касается лексического и грамматического состава текстов и сочетаемости его элементов – вплоть до букв и других графических знаков.

Новые векторы можно видеть в дальнейшем развитии социо- и психолингвистических исследований. Языковой, текстовый и корпусный материал дает сегодня богатые возможности для исследования психологических проблем личности, динамики возрастных изменений и мониторинга когнитивных процессов. Несомненным достоинством психолингвистических штудий является применение экс-

периментальных методик, которые позволяют смоделировать условия появления языкового факта, интересующего ученого. Язык и управление социальными конфликтами, изучение механизмов речи при развитии различных болезней, возрастные особенности когнитивных структур, применение технологий айтрекинга (отслеживания движения глаз) и виртуальной реальности для изучения восприятия текста, психо- и социолингвистическая характеристика пользователей социальных сетей – это только некоторые примеры проблем, решаемых на стыке социогуманитарных и естественных наук. Социальные медиа являются постоянными источниками больших данных, и особенно этот материал стал заметен с началом пандемии – в момент значительных качественных и количественных изменений в электронной коммуникации.

Не случаен и нынешний исследовательский интерес к проблемам связи языка и культуры (материальной и духовной): он наиболее явно обнаружил себя в идеях когнитивной лингвистики, лингвокультурологии, изучения «концептосферы» и «языковой картины мира» и т. д. Во всех случаях мы имеем дело с пограничными, промежуточными явлениями, которые обогащают сумму наших знаний о человеке.

Междисциплинарная лингвистика способна, таким образом, рассматривать синтетический материал – текстовый и мультимедийный, – соотносить лингвистическую и художественную графику, исследовать взаимодействие разных знаковых систем и преобразование информации при переходе из одной системы в другую. Таковы, например, специфика изменения текста в разных дискурсивных условиях (художественная проза – сценарий – кинотекст – транскрипт), проблема отображения устной спонтанной речи в письменном (в том числе художественном) тексте, специфика взаимодействия речи и жестов и др. Иначе говоря, современная лингвистика ставит задачи исследования сложных, комплексных объектов, для которого явно недостаточно резервов традиционной методологии.

На фоне таких возможностей перед исследователями возникает и базовая проблема неизменности, стабильности аппарата лингвистики в эпоху развития междисциплинар-

ных методов. Разрозненность и разнородность научно-практических векторов, обилие междисциплинарных связей ставит вопрос о возвращении к цельной и самодостаточной методологии и борьбе с размыванием лингвистических объектов. К сожалению, нередко представители смежных областей, не обладающие базовым филологическим образованием, неряшливо используют современные лингвистические ресурсы и в итоге приходят к неверным выводам, что может компрометировать лингвистические методы в глазах профессионального сообщества.

Однако в поиске новых научных парадигм и во взаимодействии со смежными сферами языкознание не утратило своего предмета. И классические источники (текст, словарь, грамматика), и современные ресурсы (корпус, электронный тезаурус, поисковая система) способствуют тому, что лингвистическое исследование становится более объективным и точным в описании различных форм речевой коммуникации.

В рубрике журнала «Филологический класс» представлены междисциплинарные возможности современной науки о языке, которые реализуются на стыке лингвистики и других гуманитарных и естественнонаучных сфер.

В статье Б. Ю. Нормана (Белорусский государственный университет, Минск) исходным является положение о содержащемся в сознании носителя языка представлении о синтаксическом «оптимуме», позволяющем наиболее результативно использовать языковые единицы для достижения коммуникативного эффекта: это совокупность интуитивных («наивных») знаний о параметрах предложения, организации его структуры, правилах соответствия лексических и грамматических значений и т. д. Отклонения от этого оптимума, проявляющиеся в повышенной частоте речевого употребления того или иного явления, могут становиться самостоятельными объектами не только научного исследования, но и пародирования в художественном тексте. Затрагиваемые лингвистические проблемы (порядок слов, немотивированный эллипсис, парцелляция и др.) перекликаются в статье с проблемами литературоведческими (пародия как жанр) и психологическими (са-

лиентность, т. е. выделимость, в том числе не основанная на оппозитивном характере).

Статья Е. Ю. Протасовой (Хельсинкский университет, Финляндия) посвящена анализу поликодовых структур, которые обнаруживаются в рекламе и вывесках, размещенных в разных регионах России. С одной стороны, современный лингвистический ландшафт должен проявлять признаки мультикультурализма – тем более в такой многонациональной стране, как Россия. С другой стороны, объемный материал исследования (несколько тысяч фотографий) показывает, что надписи, собранные в различных регионах, не отражают реального владения языками их региональных носителей. Знание собственного языка уступает влиянию государственного языка Российской Федерации, а также специфическому воздействию англоязычной (а в некоторых регионах – китайской) культуры. В то же время в материале проявляются жизненные приоритеты обычных людей, их бытовые и культурные потребности. Автор, в частности, делает вывод о том, что создание полиязыковых (гибридных) графических конструкторов может расцениваться как инструмент и показатель интернациональности и культурного разнообразия, особенно если большое количество различных этнических групп проживает рядом в одной стране.

В своем исследовании медиатекста Е. Н. Басовская (Российский государственный гуманитарный университет, Москва) и Т. А. Воронцова (Челябинский государственный университет) рассматривают соотношения фатического и информационного компонентов в профессиональных и пользовательских кинорецензиях и отзывах. Ученые задаются вопросами о возможном уменьшении «информационной плотности» высказывания за счет усиления средств контактирования с собеседником-читателем, а также о предполагаемой имитации настоящего общения средствами медиатекста. Анализ материала показывает, что авторы рецензий и отзывов решают задачу контакта с виртуальными адресатами с помощью различных средств – например, прецедентных феноменов, различных способов самопрезентации, ярких стилистических приемов (иронии и т. п.) – и осуществляют таким образом коммуникативную

настройку на «своего» собеседника. Авторы приходят к выводу, что усиление фатического компонента в тексте приводит к определенной информационной диверсификации и активизации коммуникации со «своим читателем». Такая коммуникация уподобляется разговору в поезде, когда сам процесс общения оказывается важнее информационных целей.

А. В. Колмогорова (Сибирский федеральный университет, Красноярск) проводит исследование по формализованному выявлению («детектированию») эмоциональных доминант в текстах пользователей социальных сетей. Моделирование эмотивной семантики дает возможность проведения машинного обучения на предмет автоматизации распознавания эмотивной составляющей текста. Рассматриваются вопросы о разнице в степени формализации различных эмоций (грусть, радость, гнев, удивление, стыд и др.), которая влияет на эффективность машинного обучения. В частности, на разных полюсах оказываются тексты, в которых можно детектировать проявления грусти и гнева: грусть автоматически распознается плохо, а гнев, наоборот, хорошо. В исследовании задействован краудсорсинг – для создания первоначальной базы эмотивных признаков контрольные тексты были предложены двум тысячам информантов, чье мнение сформировало основу для последующего машинного обучения. Автор формулирует социальные сценарии прототипического переживания эмоций и фактически предлагает новые технологии лексико-семантического анализа.

В статье М. В. Копотева (Хельсинкский университет, Финляндия; Высшая школа экономики, Санкт-Петербург) определяется роль корпусной лингвистики в решении различных задач современной науки. Автор дает характеристику различным видам корпусного анализа: анализ, использующий корпус (*corpus-informed analysis*), по сути выполняет иллюстративно-исследовательскую функцию; анализ, основанный на корпусе (*corpus-based analysis*) предполагает интерпретацию статистических данных; анализ, направляемый корпусом (*corpus-driven analysis*), приводит к классификации языковых явлений, проведенной как бы без участия исследователя – интерпретатора полученных больших данных. Автор делает выводы о вкладе корпусной

лингвистики в развитие языкознания, указывая на условность современного разделения языка и речи. Наиболее важным для лингвистики оказывается изучение речевой деятельности, первым делом предполагающее не моделирование правил взаимодействия языковых единиц, а описание параметров их употребления, «представляющих собой единый континуум».

Исследование, проведенное М. Ю. и Н. Ю. Мухиными (Уральский федеральный университет, Екатеринбург), реализует корпусную стилометрическую методологию в применении к лексической сочетаемости в русской прозе XIX века. В статье предлагается модель сопоставления особенностей авторской синтагматики, единицей которой является лексическая биграмма – два слова, употребленные в одном фразовом контексте. Целью исследования является систематизация оригинальных авторских сочетаний, характерных для творчества Л. Н. Толстого, Ф. М. Достоевского, А. П. Чехова, И. С. Тургенева и И. А. Гончарова. Такие примеры, как «облокотить голову» (Л. Н. Толстой), «благодарные глаза» (А. П. Чехов) или «нервический человек» (И. С. Тургенев) выявлены с помощью лексико-статистического сопоставления биграмм, в которые входит слово, часто используемое всеми авторами (*голова, глаза, человек* и т. п.). Авторскую синтагматическую модель исследователи видят как систематизированный набор таких сочетаний, которые встречаются в разных произведениях одного автора и при этом не характерны для других писателей. Статья объединяет в себе наработки в области языкознания, литературоведения, статистики и персонологии (психологии личности). В перспективе – представление этого материала в формате специализированного авторского синтагматического словаря.

Перекличка между статьями, представленными в этой рубрике, показывает, что идеи пересечения разных научных парадигм «носятся в воздухе». Рассмотрение синтетических объектов, предполагающее выход в экстралингвистические измерения, и использование больших данных для решения традиционных и новаторских задач – в этом мы видим главные междисциплинарные тренды современной лингвистики.