

УДК 811.581'37+811.581'35

Михалькова Надежда Васильевна

кандидат филологических наук, доцент,
доцент кафедры теории и практики
китайского языка
Белорусский государственный университет
иностранных языков
г. Минск, Беларусь

Nadezhda Mikhalkova

PhD in Philology, Associate Professor,
Associate Professor of the Department of Theory
and Practice of the Chinese Language
Belarusian State University of Foreign Languages
Minsk, Belarus
nadezhdakr@yandex.ru

МОДЕЛИРОВАНИЕ ПОРОЖДЕНИЯ СЛОЖНЫХ ИЕРОГЛИФОВ НА БАЗЕ ДЕТЕРМИНАТИВОВ: СЕМАНТИЧЕСКИЙ ГРАФ КИТАЙСКОЙ ПИСЬМЕННОСТИ

Данная научная статья посвящена моделированию процессов формирования сложных китайских иероглифов на основе их базовых смысловых компонентов – детерминативов. В качестве главного исследовательского инструмента использован семантический граф, который позволяет представить китайскую письменность как единую связную сеть. В статье подробно описывается механизм порождения производных знаков, выделяются ключевые семантические единицы и прослеживаются логические взаимосвязи между ними. Разработанная формальная модель успешно структурирует сложную иероглифическую систему, открывая новые перспективы для прикладных исследований в области теоретической лингвистики, компьютерного анализа текста и систем искусственного интеллекта.

Ключевые слова: *китайская письменность; семантический граф; моделирование; детерминативы; порождение.*

MODELING THE GENERATION OF COMPLEX SIGNS BASED ON THE DETERMINATIVES: SEMANTIC GRAPH OF THE CHINESE WRITING

This research article is dedicated to modeling the formation processes of complex Chinese characters based on their basic semantic components: determinatives. A semantic graph is used as the primary research tool, allowing the Chinese writing system to be represented as a single connected network. The paper describes in detail the mechanism of generating derived characters, identifies key semantic units, and traces the logical relationships between them. The developed formal model successfully structures the complex ideographic system, opening new perspectives for applied research in the fields of theoretical linguistics, computational text analysis, and artificial intelligence systems.

Key words: *the Chinese writing system; semantic graph; modeling; determinatives; generator.*

Моделирование – это способ построения семантического «скелета» китайской иероглифической письменности, позволяющего создать, с одной стороны, специфическую для китайской языковой системы «паутину смыслов» (см. семантическую паутину Т. Бернерса-Ли [1]), попеременно актуализируемых, профилируемых и «вытягиваемых» различные аспекты семантического пространства детерминативов. С другой стороны, принимая

во внимание тот факт, что набор как семантических связей, так и мотивировочных признаков китайской письменности частично изоморфен набору выявленных в других языках свойств и отношений, соединяющих смысловые точки семантического пространства словообразовательных гнезд, моделирование «работы» процесса китайского знакообразования можно квалифицировать в качестве системного общезыкового механизма. Данный механизм представляет собой комплексный процесс знакообразования, учитывающий семантический тип производной базы и производных единиц, отношения между производящей и производной базой, мотивировочные признаки выбора той или иной производящей базы, а также их вариативность. Комбинация этих этапов знакообразования в единую структуру-модель является неизбежным шагом синтеза множества последовательно полученных нами данных для перехода от линейных конструкций к ментальной «карте языкового пространства» [2] китайской письменности, позволяющей визуализировать многомерные семантические связи и ассоциативно-вербальные сети через соединяемые в единую систему участки смыслового пространства.

Применительно к знакообразованию китайской письменности идея моделирования базируется на классической дихотомии словообразования как процесса (динамическая модель) и как системы (статическая модель), развиваемой в трудах лингвистов XX в. (см. работы по словообразованию Е. С. Кубряковой [3], М. Докулила [4], П. Штекауэра [5], И. Г. Милославского [6] и др.). В частности, словообразование моделируется как двусторонний процесс кодирования и декодирования [7] в зависимости от говорящего (перспективная модель) и слушающего (ретроспективная модель). Соответственно, перспективная модель представляет собой движение от смысла к форме, а ретроспективная – от формы к смыслу. Так как словообразование – это не только система готовых единиц, но прежде всего динамическая система, обеспечивающая говорящего правилами создания новых наименований и кодом для их понимания, для слушающего производное слово выступает как динамический код, требующий расшифровки через возвращение к его истокам – производящей основе и способу ее преобразования [7]. Следовательно, моделирование знакообразования – это создание в том числе «динамического кода», где правила соединения (для говорящего) соотносятся с правилами их «расшифровки» для слушающего. Перспективный подход – это часть порождающей грамматики, алгоритма правил, позволяющего задать структуру языка и описать набор действий, с помощью которых в китайской письменности происходит построение знаков. Ретроспективный подход – это декодируемая (расшифровочная) грамматика. Пользование письменностью предполагает в китайском языке пользование частью порождающей грамматики. В результате сказанного наша задача усложняется необходимостью представления сразу двух моделей китайского знакообразования, демонстрирующих оба вектора направленности.

С технической точки зрения разрабатываемые нами модели представлены в виде *семантического графа*. Выбор *графа* в качестве основного инструмента отображения моделей выявленных нами семантических связей и мотивировочных признаков выбора детерминативов китайской письменности обусловлен принципом представления данных в этой абстрактной структуре дискретной математики. По аналогии с графами знаний (knowledge graph) в Google, это способ организации данных, который описывает объекты и связи между ними в виде сети. В отличие от обычной базы данных как плоской таблицы (например, Excel), граф – это многомерная паутина, где узлы (точки) – это сущности, а линии-связи между ними – отношения. Граф позволяет представить все семантические связи китайской письменности в виде «паутины смыслов». Созданный нами семантический граф пространства китайской письменности статически «моделирует» действующую на сегодня иероглифическую систему. Это не исключает возможности прогнозирования ее развития, т.е. создания новых единиц по тем же механизмам, «знакомым» именуемым субъектам – носителям китайского языка, и декодирования на основе установленных семантических связей и мотивировочных признаков смыслов как существующих, так и вводимых в систему новых знаков. Теория графов – один из фундаментальных разделов дискретной математики, которая является основой программирования. Она позволит не только представить все семантические связи китайской письменности в виде «скелета смыслов», но и впоследствии (при необходимости) переложить представленные нами семантические графы на язык программирования. Кроме того, как утверждает Ю. Д. Апресян, «язык семантических графов является, по предположению, универсальным – единым для всех человеческих языков ... создание такого семантического языка продолжает оставаться конечной задачей лингвистической семантики» [8, с. 21], без которой, по мнению Г. В. Колшанского, «не может быть теории языка» [9, с. 6].

Следует отметить, что термин «семантический граф» не раз употреблялся лингвистами для формализации тех или иных смыслов. В частности, Франсуа Растье [10] используется концепция семантического графа для формализации смысла текста, В. Б. Касевичем семантические графы применялись для описания структур предложения [11]. В определенном смысле семантическими графами можно считать экзистенциальные графы Ч. Пирса [12], считавшего их логикой будущего; семантические сети как модель человеческой памяти Р. Куиллиана [13]; стеммы как визуальные схемы «распаковки» линейной структуры предложения Л. Теньера [14], семиотический квадрат Греймаса – визуальный граф того, как одно понятие порождает целую систему смыслов [15]; графовые модели Б. Потье для компонентного анализа семантики слова [16]; концептуальные графы Дж. Сова, принятые сегодня в качестве стандарта в области искусственного интеллекта [17]; модель «смысл-текст» И. Мельчука, Ю. Д. Апресяна и А. Жолковского [18] и др. Благодаря указанным выше исследованиям

семантических графов был совершен прорыв в разработке систем искусственного интеллекта, позволивший перейти от нейросетей, которые предсказывают следующий элемент на основе статистики, к ИИ на базе графов, оперирующему структурированными знаниями и возможностью самообучения (см. работы Р. Пиотровского по инженерной лингвистике, лингвистической синергетике, теории лингвистических автоматов [19; 20], цепи А. А. Маркова, моделирующие вероятность появления следующего слова или фразы на основе имеющихся контекстов [21], стохастичность текста Е. В. Шелестюк [22], статистические законы распределения слов Ципфа–Мандельброта [23], язык СЕМСИНТ А. В. Зубова [24], создание теории автоматического порождения архитектуры знаний (ТАПАЗ-2) школы комбинаторной семантики А. Н. Гордея, интегральное моделирование индивидуального стиля В. С. Андреева [25] и др.). Подобный междисциплинарный подход привел к возможности создания универсального метаязыка, связывающего данные из разных областей, в том числе лингвистики, в единую интеллектуальную сеть (см. Естественный Семантический Метаязык (NSM) как «алфавит человеческой мысли» А. Вежбицкой).

В данном исследовании нами представлены два типа семантических графов, репрезентирующих, соответственно, *перспективную* (от детерминативов к сложным иероглифам, их включающим) и *ретроспективную* (от сложных иероглифов к детерминативам в их составе) модели выбора детерминативов китайской письменности. Отметим также, что «любое структурное описание объекта, не вызывающее возражений, поскольку является общей чертой науки как таковой», тем не менее неизбежно приводит к определенному его упрощению и «д о о р г а н и з о в а н н о с т и», т.е. объект становится «более жестко организованным, чем это имеет место на самом деле» [26, с. 5].

Методика построения семантических графов соответствует четырем шагам их конструирования, принятым в дискретной математике и последовательно осуществляемым в наших моделях.

1. Определение множества вершин (точек).
2. Определение множества ребер (связей).
3. Выбор типа связи (направленность).
4. Присвоение весов или типов (меток ребра).

В соответствии с этими шагами для построения семантических графов китайской иероглифической письменности нами на первом и втором шагах конструирования были определены вершины (точки) и множества ребер (связей). Поскольку выбор детерминатива представляет собой движение или от конечного результата (производного сложного иероглифа) к семантическому компоненту (производящей базе – детерминативу) – ретроспективная модель, или от детерминатива к производному сложному иероглифу (перспективная модель), разработка семантических графов началась с установления тех точек и ребер-путей, соединяющих области сложных

иероглифов и детерминативы, которые в итоге включаются в их состав (рис. 1). В нашем исследовании мы конструируем двунаправленный семантический граф китайской письменности, который представляет соответственно две модели одновременно: ретроспективную, т.е. предполагающую движение от сложных иероглифов (производных единиц) к детерминативам (производящим базам), и проспективную – динамическую модель, которая показывает обратный процесс.

Построенный нами семантический граф это не что иное, как одновременно *статическая* (вектор движения сверху вниз) и *прогнозируемая модель* (вектор движения снизу вверх), т.е. механизм, который описывает и то, что уже есть, и то, как может быть создано новое. Это своеобразный алгоритм, по которому человек создает вводимые в систему языка знаки, «проходя» от цели (обозначить тот или иной объект, статическое или динамическое свойство и др.) к форме, с помощью которой она выражается.

Представим далее ретроспективную модель семантического графа.

Первая линия ретроспективной модели семантического графа – это совокупность точек А1-А6, репрезентирующих соответственно 6 типов конечного результата:

А1 – сложные иероглифы, обозначающие субъект,

А2 – сложные иероглифы, обозначающие объект О',

А3 – сложные иероглифы, обозначающие вещество,

А4 – сложные иероглифы, обозначающие динамическое свойство (действие),

А5 – сложные иероглифы, обозначающие статическое свойство (признак),

А6 – сложные иероглифы, обозначающие абстрактные сущности.

Вторая линия графа – это типы семантических связей, которые связывают детерминативы и сложные иероглифические знаки: гиперо-гипонимия, холо-меронимия, метонимия, метафора, метафтонимия, репрезентируемые соответственно точками В1-В5.

Третья линия семантического графа представляет собой следующий этап выбора детерминатива – выбор мотивировочных признаков. Это точки С1-С8, выражающие *материал* (С1), *субъект* (С2), *форма* (С3), *объект* (С4), *локация* (С5), *функция* (С6), *размер* (С7), *инструмент* (С8).

Четвертая линия семантического графа – это семантический тип детерминатива, выбор которого соответствует тому или иному типу семантических связей (вторая линия В1-В5) и мотивировочных признаков (третья линия С1-С8).

Семантический тип детерминатива соответствует нашей семантической классификации и включает: детерминативы-знаки природных объектов (D1), детерминативы-знаки жилищ и их частей (D2), детерминативы-соматизмы (D3), детерминативы-фитонимы (D4), детерминативы-зоонимы (D5), детерминативы-знаки лиц (D6), детерминативы-знаки абстрактных сущностей (D7), детерминативы-знаки веществ (D8).

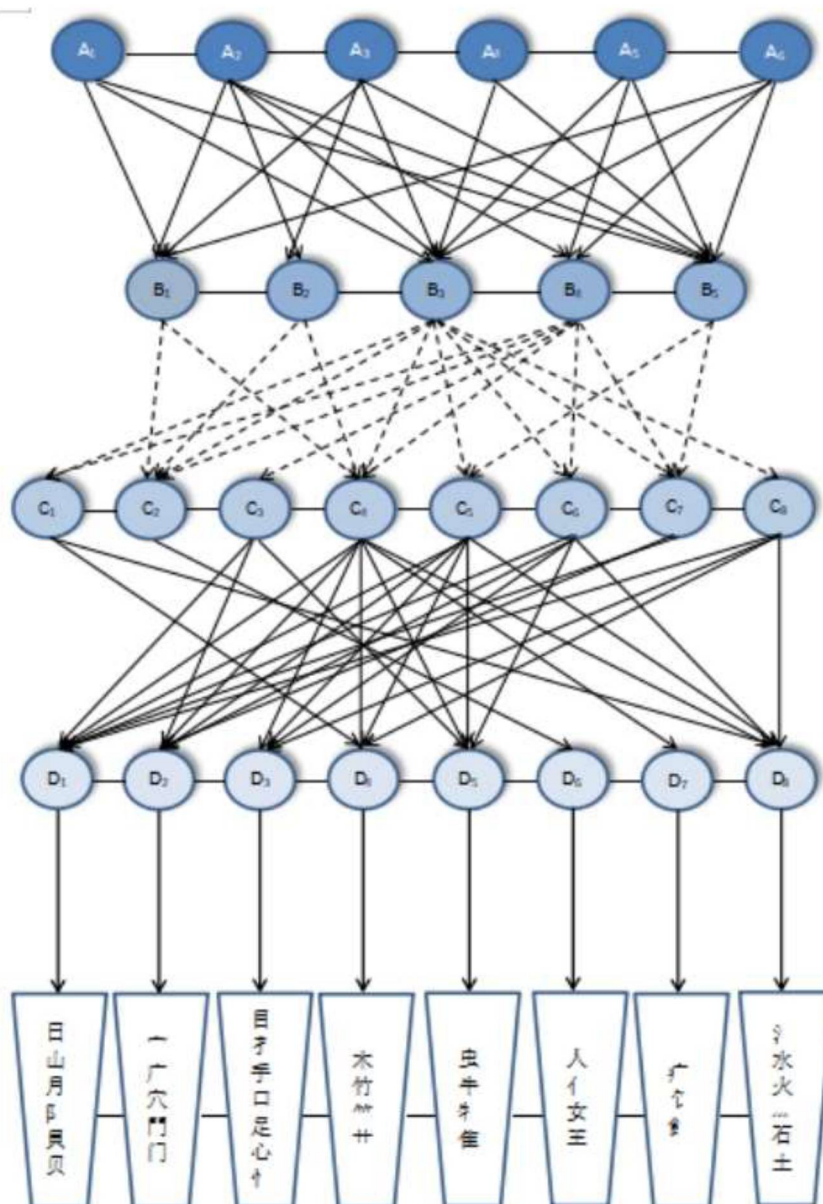


Рис. 1. Семантический граф китайской иероглифической письменности (ретроспективная модель)

Ребра семантического графа выстроены в соответствии с установленными нами ранее возможностями системы китайской письменности формировать те или иные семантические связи. Ребра семантического графа имеют свою специфику, заключающуюся в вариативности выбора мотивировочных признаков, что обусловило использование не прямых, а прерывистых линий, соединяющих точки B1-B5 и ряд точек C1-C8 (шаг третий).

Наконец, четвертый шаг построения семантического графа по аналогии с шагами в дискретной математике включает присвоение весов или типов (меток ребер). В конструируемом нами семантическом графе китайской иероглифической письменности метками ребер являются числовые данные *семантических связей*, возникающих между соответствующими

соединениями линий A1-A6 и B1-B5, числовые данные *мотивировочных признаков* выбора (между линиями B1-B5 и C1-C8), *детерминативы*, которые отвечают тем или иным «требованиям».

Таким образом, построенный нами семантический граф китайской письменности является *ориентированным* (ребра имеют определенное направление – стрелки), *взвешенным* (ребрам присваивается число-вес, коэффициенты), *связным* (из вершины можно добраться в другую вершину) *мультиграфом* (разрешены несколько ребер от одной вершины).

Рассмотрим на примере процедуру «работы» семантического графа китайской письменности. Статическая модель семантического графа китайской письменности осуществляет процедуру выбора детерминатива для того или иного требуемого обозначения, т.е. отвечать на вопрос: с помощью каких детерминативов обозначается то или иное явление, объект, действие, абстрактная сущность и др. Допустим, что именуемому субъекту необходимо выбрать детерминатив для репрезентации в сложном знаке китайской письменности динамического свойства – действия «защищать». Вес ребра по линии метонимии является наиболее высоким (см. рис. 1). Следовательно, с большой вероятностью данные обозначения будут образовываться путем метонимических переносов. Далее необходимо выбрать мотивировочный признак, который может быть ведущим при метонимии. Весовые параметры ребер B3-C5 и B3-C8 позволяют заключить, что чаще всего действие обозначается через мотивировочный признак *локация* (место) или через мотивировочный признак *объект* как инструмент, с помощью которого оно выполняется. Соответственно, движемся далее или по ребру B3-C5, или по ребру B3-C8 (вариативность выбора). Согласно полученным нами ранее индексам вариативности детерминативов с большой вероятностью необходимо произвести далее выбор или из детерминативов-обозначений жилищ и их частей (если движемся по ребру B3-C5 и выбираем мотивировочный признак *локация*), или из детерминативов-соматизмов (если движемся по ребру B3-C8 и выбираем мотивировочный признак *инструмент*). В результате китайская иероглифическая письменность располагает следующими детерминативами-обозначениями жилищ и их частей для построения сложного знака «защищать»: 广 ‘навес’ и 宀 ‘крыша’, среди детерминативов-соматизмов – детерминатив 扌 ‘рука’. Смотрим знакообразовательные гнезда с данными детерминативами. Находим искомые сложные знаки с детерминативом 宀 ‘крыша’ (守 ‘защищать (чиновника)’, 寨 ‘охранять’), с детерминативом 广 ‘навес’ – 庇 ‘защищать’, а также с детерминативом 扌 ‘рука’ – 捍 ‘отражать, защищать’, 护 ‘защищать, беречь’, 扞 ‘защищать, отражать врага’.

Аналогично представленному выше примеру будет «работать» семантический граф в обратную сторону – вектор движения от производящей базы (детерминатива) к той производной единице (сложному знаку), который потенциально может образоваться.

Полученный семантический граф лег в основу разработки компьютерной программы **DetGen-Model**, позволяющей отобразить механизм порождения исследованных сложных иероглифов относительно включения детерминативов различного типа (рис. 2).



Рис. 2. Интерфейс компьютерной программы DetGen-Model

Проспективная модель представлена в виде возможности ввода детерминатива и получения в выводе всего семантического спектра сложных иероглифов, которые могут содержать введенный детерминатив (на основе исследованной части китайской письменности) (рис. 3).



Рис. 3. Пример работы проспективной модели компьютерной программы DetGen-Model

Ретроспективный режим работы программы DetGen-Model позволяет при введенном искомом понятии на русском языке отобразить при выводе те детерминативы, которые могут составлять сложный иероглиф в китайском языке. При этом программа может согласно графу семантических связей предложить все варианты детерминативов. В качестве примера иероглифов будут представлены только те, которые содержатся в исследованной части китайской письменности и находятся в базе данных программы (рис. 4).



Рис. 4. Пример работы ретроспективной модели компьютерной программы DetGen-Model

Таким образом, семантический граф наглядно показывает как ведущие линии связей, так и лакунарные области, и в перспективе может являться смысловым атласом знакообразования китайской письменности с возможностью математического доказательства, какая модель «умирает», а какая становится доминирующей. Будучи фундаментом статической и динамической моделей знакообразования исследованной части китайской письменности, это эффективный способ не только «объяснить машине», как выстраивается логика выбора детерминативов сложных иероглифических знаков, но и развить соответствующие навыки у обучающихся китайскому языку. Кроме того, построенный нами семантический граф может явиться одной из основ для развиваемого сегодня нового направления High-Tech лингвистики, которое переносит изучение языка в область точных данных, алгоритмов и искусственного интеллекта, отвечая на вопрос о том, как сделать так, чтобы машина работала с языком так же эффективно, как это делает человек.

ЛИТЕРАТУРА

1. The Semantic Web / T. Berners-Lee, J. Hendler, O. Lassila // Scientific American Magazine. – 2001. – Vol. 284, № 5. – P. 34–43.
2. *Haspelmath, M.* The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison / M. Haspelmath // The new psychology of language. – NJ : Lawrence Erlbaum, 2000. – Vol. II. Mahwah / Tomasello Michael (ed.). – P. 211–242.
3. *Кубрякова, Е. С.* Язык и знание: на пути получения знаний о языке: Части речи с когнитивной точки зрения. Роль языка в познании мира / Е. С. Кубрякова. – М. : Языки славян. культуры, 2004. – 560 с.
4. *Dokulil, M.* Tvoreni slov v cestine / M. Dokulil. – Praha : Czechoslovakia Nakladatelství Československé Akademie Věd, 1967. – Vol. 2: Odvozování podstatných jmen. – 779 s.
5. *Štekauer, P.* An Onomasiological Theory of English Word-Formation / P. Štekauer. – John Benjamins Publishing Company, 1998. – 202 p.
6. *Милославский, И. Г.* Вопросы словообразовательного синтеза / И. Г. Милославский. – М. : Изд-во МГУ, 1980. – 296 с.
7. *Кубрякова, Е. С.* Типы языковых значений. Семантика производного слова / Е. С. Кубрякова. – М. : Наука, 1981. – 198 с.
8. *Апресян, Ю. Д.* Избранные труды : в 2 т. / Ю. Д. Апресян. – 2-е изд., испр. и доп. – М. : Восточная литература, 1995. – Т. 2 : Интегральное описание языка и системная лексикография. – 767 с.
9. *Колшанский, Г. В.* Некоторые вопросы семантики языка в гносеологическом аспекте / Г. В. Колшанский // Принципы и методы семантических исследований / редкол.: В. Н. Ярцева (отв. ред.) [и др.]. – М. : Наука, 1976. – С. 5–31.
10. *Растье, Ф.* Интерпретирующая семантика / Ф. Растье ; пер. с фр., примеч., предм.-имен. указ. А. Е. Бочкарева. – Н. Новгород : Деком, 2001. – 367 с.
11. *Касевич, В. Б.* Семантика. Синтаксис. Морфология / В. Б. Касевич. – М. : Наука, Главная редакция восточной литературы, 1988. – 311 с.
12. *Roberts, Don D.* The existential graphs of Charles S. Peirce / Don D. Roberts. – The Hague : Mouton, 1973. – 176 p.
13. *Quillian, R.* Semantic memory / R. Quillian. – Pittsburgh : Carnegie Institute of Technology, 1966. – 222 p.
14. *Теньер, Л.* Основы структурного синтаксиса / Л. Теньер ; пер. с фр. И. М. Богуславского [и др.] ; вступ. ст. [с. 5–21], общ. ред. [и коммент.] В. Г. Гака. – М. : Прогресс, 1988. – 653 с.
15. *Греймас, А.-Ж.* Структурная семантика. Поиск метода / А.-Ж. Греймас ; пер. Л. Зиминой. – М. : Академ. Проект, 2004. – 367 с.
16. *Степанов, Ю. С.* Методы и принципы современной лингвистики / Ю. С. Степанов. – М. : Наука, 1975. – 310 с.

17. Sowa, J. F. Conceptual Structures: Information Processing in Mind and Machine / J. F. Sowa. – Reading, Mass. : Addison-Wesley, 1984. – 481 p.
18. Апресян, Ю. Д. Восемь словарных статей толково-комбинаторного словаря русского языка / Ю. Д. Апресян, А. К. Жолковский, И. А. Мельчук ; Акад. наук СССР, Ин-т рус. яз. Проблем. группа по эксперим. и приклад. лингвистике. – М., 1970. – 31 с.
19. Пиотровский, Р. Г. Инженерная лингвистика и теория языка / Р. Г. Пиотровский. – Л. : Наука. Ленингр. отд-ние, 1979. – 112 с.
20. Пиотровский, Р. Г. Лингвистический автомат и его речемыслительное обоснование / Р. Г. Пиотровский. – Мн. : Мин. гос. лингвист. ун-т, 1999. – 195 с.
21. Shannon, C. E. A Mathematical Theory of Communication / C. E. Shannon // Bell System Technical Journal. – 1948. – Vol. 27. – P. 379–423.
22. Шелестюк, Е. В. Стохастичность текста : сущность, способы исследования, феноменология / Е. В. Шелестюк, Е. А. Щетинкина // Вестник Челябинского государственного университета. Филологические науки. Вып. 124. – 2021. – №4 (450). – С. 173–184.
23. Mandelbrot, B. Information theory and psycholinguistics : a theory of word frequencies / B. Mandelbrot // Readings in mathematical social science. – Cambridge, Mass. [u.a.] : M.I.T. Press, 1968. – P. 350–368.
24. Зубов, А. В. Вероятно-алгоритмическая модель порождения текста : Семантико-синтактический аспект : дис. ... д-ра филол. наук : 10.02.21 / Зубов Александр Васильевич ; МГПИИЯ. – Мн., 1985. – 399 л.
25. Андреев, В. С. Моделирование индивидуального стиля на основе лингвистических характеристик : монография / В. С. Андреев. – М. : Флинта : Наука, 2012. – 247 с.
26. Лотман, Ю. М. Динамическая модель семиотической системы / Ю. М. Лотман. – М. : ИРЯ АН СССР, 1974. – 28 с.

Поступила в редакцию 12.06.2026