

QO‘QON DAVLAT
PEDAGOGIKA INSTITUTI
ILMIY XABARLARI
(2025-yil 3-son)



FILOLOGIYA
PHILOLOGY

RUS VA INGLIZ TILLARIDAGI LEKSIK NUKLEOLOGIYANING QIYOSIY TAHLILI

Tagayeva Asal Rustamovna
Navoiy davlat universiteti
asal.tgv99@mail.ru

Annotatsiya. Ushbu maqolada rus va ingliz tillari leksik nukleologiyasining korpus lingvistikasi ma’lumotlari hamda so‘z qo‘llanishlarining statistik tadqiqiga asoslangan keng qamrovli qiyosiy tahlili keltirilgan. Muallif “yadroviy leksika” tushunchasini nutqda asosiy kommunikativ vazifani bajaradigan eng barqaror va universal birliklar majmuasi sifatida yoritib beradi. Shu bilan birga, umumiy tipologik xususiyatlarni (yordamchi so‘zlarning ustunligi, asosiy antropotsentrik tushunchalar) aniqlaydi va grammatik tuzilish bilan bog‘liq farqlarni (ingliz tilining analitikligi va rus tilining sintetikligi) ko‘rsatib o‘tadi. Asosiy leksemalarning semantik rivojlanish mexanizmlari va ularning ko‘p ma’nodlilik, shuningdek, grammatikalashgan elementlarning (artikllar, yordamchi fe‘llar, yuklamalar) roli alohida e’tibor bilan o‘rganilgan.

Kalit so‘zlar: leksik nukleologiya, qiyosiy tahlil, rus tili, ingliz tili, asosiy leksika, korpus lingvistikasi, semantik yadro, chastotali lug‘at.

КОМПАРАТИВНЫЙ АНАЛИЗ ЛЕКСИЧЕСКОЙ НУКЛЕОЛОГИИ В РУССКОМ И АНГЛИЙСКОМ ЯЗЫКАХ

Аннотация. В статье представлен комплексный компаративный анализ лексической нуклеологии русского и английского языков, основанный на данных корпусной лингвистики и статистического исследования частотных словоупотреблений. Автор раскрывает понятие «ядерной лексики» как совокупности наиболее устойчивых и универсальных единиц, обеспечивающих основную коммуникативную функцию в речи, и выявляет общие типологические признаки (преобладание служебных слов, базовые антропоцентричные концепты) при одновременных различиях, связанных с морфологией и грамматической структурой (анализируем английский язык и синтетизм русского). Особое внимание уделено механизмам семантической эволюции ключевых лексем и их полисемии, а также роли грамматикализованных элементов (артикли, вспомогательные глаголы, частицы).

Ключевые слова: лексическая нуклеология, сопоставительный анализ, русский язык, английский язык, базовая лексика, корпусная лингвистика, семантическое ядро, частотный словарь.

COMPARATIVE ANALYSIS OF LEXICAL NUCLEOLOGY IN RUSSIAN AND ENGLISH

Abstract. This article presents a comprehensive comparative analysis of lexical nucleology in Russian and English, based on corpus linguistics data and a statistical investigation of frequent word usages. The author introduces the concept of “nuclear vocabulary” as a set of the most stable and universal units that ensure the primary communicative function in speech, and identifies common typological features (the predominance of function words, basic anthropocentric concepts) alongside differences related to morphology and grammatical structure (the analytic nature of English versus the synthetic nature of Russian). Particular attention is paid to the mechanisms of semantic evolution of key lexemes and their polysemy, as well as to the role of grammaticalized elements (articles, auxiliary verbs, particles).

Keywords: lexical nucleology, comparative analysis, Russian language, English language, basic vocabulary, corpus linguistics, semantic core, frequency dictionary.

Введение. Лексическая нуклеология представляет собой направление в лексикологии, посвящённое систематическому исследованию «ядерной лексики» — совокупности наиболее фундаментальных, частотных и устойчивых единиц словарного состава языка. Согласно гипотезе лексической стабильности (сродни идеям глоттохронологии), в каждом языке сохраняется базовый корпус ключевых понятий, который мало подвержен историческим изменениям и формирует ядро лексической системы [8; с. 517]. Сопоставительный анализ такого ядра в русском и английском языках позволяет выявить общие типологические признаки (повторяющиеся семантические категории: части тела, природные явления, родственные отношения, служебные слова), а также специфические особенности, связанные с морфологическим оформлением, функциональной нагрузкой, степенью включённости в текст и т.д. Изучение эволюции «ядерных» единиц проясняет механизмы семантического расширения, появление новых значений и их дальнейшую трансформацию, что представляет ценность как для теории лексикологии и типологии, так и для прикладных задач: совершенствования двуязычных словарей, методик преподавания, переводческой практики и интеграции результатов в компьютерную лингвистику.

Методы. Исследование базируется на сравнительно-сопоставительном (компаративном) методе, предполагающем соотнесение базовых лексических единиц русского и английского языков по семантическим полям и функциональным категориям. Такой подход обеспечивает выявление эквивалентов (совпадающих по смыслу ядерных

понятий, «вода», «земля», личные местоимения) и лакун (единиц, присутствующих в одном языке и отсутствующих в другом). Компаративный анализ охватывает как отдельные лексемы, так и их группировки по частям речи либо семантическим разрядам, что позволяет формировать целостную картину организации «ядерного» слоя лексики двух языков.

Для определения ядерных лексем основным критерием выступила их частотность, установленная методом статистического анализа словоупотребления. Списки наиболее часто употребляемых слов в русском и английском языках были сопоставлены, и оценена доля таких единиц в общем объёме текста. Относительно небольшой набор лексем (порядка 1000 слов) охватывает значительную часть текстовой выборки (до 75–80% в английском и сопоставимые показатели в русском). Это согласуется с законом Ципфа, указывающим на особую роль узкого слоя высокочастотных единиц в языковой системе.

Эмпирической базой для анализа послужили Национальный корпус русского языка (НКРЯ) [11] и Британский национальный корпус (BNC) [6], а также опубликованные на их основе частотные словари и списки (частотный словарь русского языка под ред. О.Н.Ляшевской и С.А.Шарова) [10]. Этапы корпусного анализа включали: 1. извлечение топ-1000 лемм из каждого корпуса, 2. классификацию их по частям речи и семантическим классам, 3. статистическое сопоставление (доля местоимений, наличие/отсутствие артиклей и т. п.), 4. качественный анализ примеров употребления для выявления спектра значений и функций. Итоговые количественные данные представлены в виде таблиц и диаграмм, демонстрируя структурные сходства и различия в «ядерной» лексике русского и английского языков.

Результаты. *Классификация нуклеологических единиц.* Ядро лексики и в русском, и в английском языках характеризуется преобладанием слов грамматически служебного и базового семантического характера. К нуклеологическим единицам, согласно нашим критериям, отнесены прежде всего: а. местоимения (личные, указательные и др.), б. основные числительные (особенно «one, two, три, два» и т.д.), с. простейшие глаголы общей семантики (глаголы движения: «идти», «go»; существования: «быть», «be»; владения: «иметь», «have»; базовые действия: «дать», «give» и т.п.), d. простейшие существительные, обозначающие ключевые объекты и понятия (*человек, time время, год, year, рука hand, дом house, вода, water* и др.), е. наиболее общие прилагательные (большой, big; новый, new; хороший, good и т.д.), а также f. широкий класс служебных слов – предлоги, союзы, частицы, артикли (в английском).

Для наглядности в таблице 1 представлены десять наиболее частотных лексем русского и английского языков (по данным корпусов), с указанием их грамматической категории и примерного значения:

Таблица 1. Десять наиболее частотных слов русского и английского языков (леммы) и их категория

№	Русский (лемма)	Категория	Английский (лемма)	Катгория
1	и	союз (сочинительные)	the	определённый артикль
2	в	предлог (места/направления)	of	предлог (род. падеж, принадлежность)
3	не	частица (отрицание)	and	союз (координативный)
4	на	предлог (места/направления)	to	частица/предлог (инфинитив, направление)
5	я	местоимение (1 л., ед.ч.)	a	неопределённый артикль
6	он	местоимение (3 л., ед.ч., м.р.)	in	предлог (места)
7	быть	глагол (связка «быть»)	to be (be)	глагол (связка «быть»)
8	что	местоимение/союз (вопросит./подч.)	is (be)	глагол (форма «быть», 3 л.)
9	с («с/со»)	предлог (совместность/откуда)	I	местоимение (1 л., ед.ч.)
10	это	местоимение/указательное	it	местоимение (3 л., ед.ч.)

Анализ топ-10 лексем показывает явное преобладание служебных единиц в обоих языках. В русском первыми позициями выступают союзы (и), предлоги (в, на, с), отрицательная частица (не), местоимения (я, он, что, это) и глагол-связка *быть*. В английском соответствующие места занимают артикли (the, a), предлоги (of, in), союз (and), частица инфинитива (to), формы глагола to be (is) и местоимения (I, it). Такая структура указывает на грамматические различия: в английском артикли формируют отдельную высокочастотную категорию, тогда как в русском их функцию частично выполняют указательные местоимения (этот, то). В обоих языках в абсолютное ядро входят личные местоимения (я — I, он — he, it — оно/это), предлоги, союз *and* — *и* и глагол-связка *быть* — to be (при различиях в синтаксическом употреблении).

Помимо служебных слов, в расширенное ядро (топ-100, топ-1000) входят многочисленные *общеупотребительные знаменательные слова*. Можно выделить несколько семантических классов, образующих ядро лексики обоих языков:

природные объекты и явления: *sun* – *солнце*, *moon* – *луна*, *water* – *вода*, *fire* – *огонь*, *earth* – *земля* и др. Эти слова обозначают базовые реалии окружающего мира и восходят к древнейшим слоям лексики;

части тела: *hand* – *рука*, *eye* – *глаз*, *head* – *голова*, *heart* – *сердце* и т.п. Являются фундаментальными понятиями, непосредственно связанными с человеческим опытом;

близкие родственные отношения: *mother* – мать, *father* – отец, *son* – сын, дочь – *daughter*, брат – *brother*, сестра – *sister* и т.д. Родственные термины обычно относятся к ядерной лексике и высокочастотны в базовом общении;

пространственно-временные понятия: *day* – день, *night* – ночь, *год* – *year*, время – *time*, место – *place*, дома – *home* (как «домой») и др. Без этих слов невозможно выражение базовых категорий времени и пространства;

основные действия и движения: глаголы типа *идти* – *go*, *бежать* – *run*, *дать* – *give*, *взять* – *take*, *делать* – *do*, *сделать* – *make*, *видеть* – *see*, *слышать* – *hear* и др. Они принадлежат к ядру потому, что описывают простейшие повседневные действия и ощущения;

качественные характеристики: *большой* – *big*, *маленький* – *small*, *хороший* – *good*, *плохой* – *bad*, *новый* – *new*, *старый* – *old* и некоторые другие базовые прилагательные, обозначающие простые и универсальные признаки.

Отдельно следует упомянуть разряд слов, специфичных для каждого языка, которые тоже могут входить в ядро по частотности или функции. Например, в русском языке среди топ-100 частотных слов присутствует частица-модальность «бы» (используется в условных конструкциях и желательных наклонениях), а в английском – вспомогательный глагол *will* (маркер будущего времени) или частотное наречие *not* (в форме частички *n't*). Эти слова выполняют сугубо грамматическую роль, однако их высокая частотность делает их частью функционального ядра лексики. Таким образом, классификация нуклеологических единиц включает как *семантическое ядро* (набор универсальных знаменательных слов), так и *функциональное ядро* (набор служебных, грамматических слов), обеспечивающих структуру высказываний.

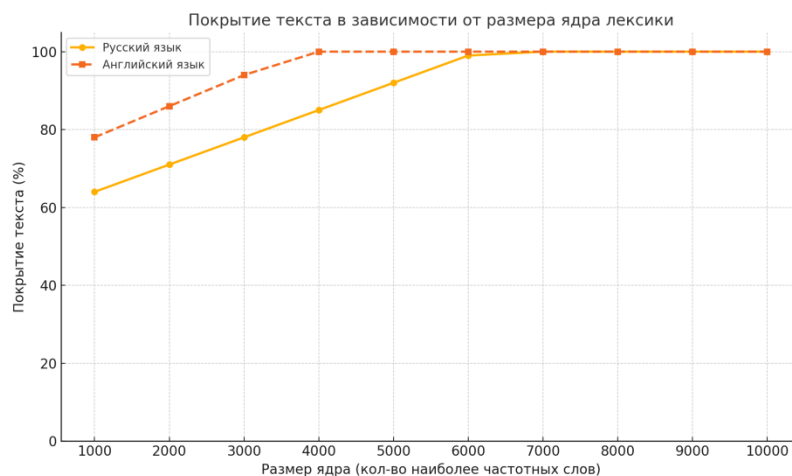
Результаты сопоставления показывают, что ядро лексики русского и английского языков в количественном отношении сходно, но качественно имеет отличия вследствие типологических характеристик языков. Оба языка подчиняются закону частотного распределения (закону Ципфа): несколько сотен наиболее частых слов покрывают львиную долю текстов. При этом доля служебных слов среди топ-100 английского выше, чем в русском, из-за наличия артиклей и аналитических форм. В русском, напротив, значительную часть высокочастотного списка составляют разные формы одних и тех же знаменательных лексем, обусловленные богатой морфологией. Лемма «человек» (англ. *person/man*) встречается во множестве форм (человек, люди, человеку, людей и т.д.), каждая из которых отдельно может не войти в топ-100 по частоте, но совокупная частотность леммы весьма высока. В английском же эквивалент *man/person* не склоняется (за исключением формы *men*), и потому основная форма *man* имеет более высокое место в списке. Для учёта этого фактора мы производили анализ по леммам (нормализованным словоформам).

Анализ показал, что в английском топ-1000 около 55–60% лемм составляют грамматические слова (артикли, местоимения, предлоги, союзы, модальные и

вспомогательные глаголы), тогда как в русском доля служебных слов несколько ниже около ~45%. Это объясняется тем, что русский язык выражает многие грамматические смыслы морфологически (через приставки, флексии), а не отдельными служебными словами. Категория определённости в русском отсутствует в виде артиклей; время и залог зачастую выражаются морфологически (суффиксами, окончанием глагола), поэтому в русском нет эквивалента английским вспомогательным *do*, *will* или артиклю *the*. Тем не менее, ряд служебных категорий присутствует в обоих языках (предлоги, союзы) и их представители входят в ядро. В частности, пространственные предлоги *в*, *на*, *с* в русском соответствуют *in*, *on*, *with/from* в английском; сочинительный союз *и* соответствует *and*; подчинительный *что* соответствует *that* (в определительной и изъяснительной функции). Отличие в том, что английский *that* разграничивает указательное местоимение и союз (хотя графически совпадает), тогда как русское *что* выполняет обе роли без формального различия – отсюда чрезвычайно высокая частотность слова «что», покрывающего две функции.

Измерение покрываемости текста ядерной лексикой подтвердило ожидаемую тенденцию: небольшое ядро обеспечивает коммуникацию в большинстве случаев. В английском языке 1000 наиболее частых слов покрывают до ~75–80% типичного текста, в русском несколько меньше – порядка 64%, что частично обусловлено учётом разных словоформ. Увеличение размера ядра даёт убывающую отдачу: добавление следующих 1000 слов увеличивает покрытие текста всего на ~8–10% и т.д. (см. рис. 1). График частотности (Рис. 1) отображает резкое падение частот после первых позиций: самая частая лексема (англ. *the*, рус. *и*) встречается гораздо чаще, чем 50-я. Это отражает универсальную закономерность: ядро лексики весьма ограничено по составу, а периферия огромна.

Рис. 1. Распределение частотности лексических единиц (ранг 1–1000) в русском и английском: доля охвата текста.



Под *семантической трансформацией* в контексте ядра лексики мы понимаем развитие значений и смысловых оттенков у базовых слов по мере их использования в различных контекстах. Выявлено, что ядерные лексемы, как правило, обладают широкой

полисемией и способностью к переносному употреблению. Русское слово «голова» изначально обозначает часть тела, но в ядерном словаре оно также выступает в значении «ум, рассудок» («светлая голова» – умный человек) и входит во множество фразеологизмов. Английское *head* демонстрирует сходную полисемию (глава организации, вершина чего-либо, ум, пена на пиве и т.д.). Таким образом, *ядерное значение* (прямое номинативное) служит основой, от которой расходятся переносные значения. По меткому выражению В.В.Виноградова, «прямое номинативное значение слова является непосредственной связью с предметом или явлением и представляет собой «фундамент всех других значений» данного слова» [9; с. 12]. Наш анализ подтверждает эту мысль: наиболее частотные слова обычно имеют одно или несколько центральных значений, наиболее устойчивых и частых, от которых исторически и контекстуально развиты остальные, более специализированные или образные значения. Ядерные лексемы формируют семантические гнёзда с радиальной структурой: ядро значения – ближние переносные значения – дальние, менее прототипичные.

Ядерные лексемы в ряде случаев переходят от собственно лексического значения к грамматическим или служебным функциям. Так, «быть» в русском и *to be* в английском сохраняют семантику «существовать», но одновременно функционируют как связки и вспомогательные элементы. Аналогично указательное местоимение *that* в английском, помимо демонстративного значения, служит союзом; русское «что» совмещает вопросительную и союзную роли. Эти примеры демонстрируют тесное переплетение лексического ядра с грамматическим строем, поскольку наиболее частотные слова нередко грамматикализуются. При этом историческая эволюция ядра происходит медленно: базовые лексемы сохраняют свои центральные значения и лишь постепенно расширяют функциональный спектр (ср. местоимение *it* в безличных конструкциях современного английского).

Обсуждение. Полученные данные подтверждают существование лексических универсалий: в любом языке присутствуют средства выражения базовых антропоцентричных значений (по Д.Лайонзу) [4], что согласуется с тезисом об универсальных категориях мышления. При этом системные различия существенно влияют на функционирование нуклеологических единиц. Синтетический русский опирается на развитое словоизменение и в меньшей мере на служебные слова; аналитический английский вынес грамматические показатели в отдельные лексемы (артикли, модальные глаголы, вспомогательные элементы), что повышает их частотность и полифункциональность. Так, категория определённости/неопределённости представлена в английском артиклями *the*, *a*, вошедшими в ядро по частоте, тогда как в русском аналогичная функция не оформлена отдельным служебным классом. Культурные различия (например, «рубль», «Москва» в русском и “Mr”, “Mrs” в английском) отражают специфику корпусов и не относятся к универсальному ядру,

поэтому для выявления базовой лексики их следует исключать как периферийные элементы.

Итоги сопоставления хорошо согласуются с положениями ведущих лингвистических школ. Так, В.В.Виноградов, вводя понятие прямого (номинативного) и производных значений, фактически заложил основу концепции ядерного (инвариантного) значения: каждая лексема в ядре сохраняет неизменный семантический стержень, вокруг которого группируются производные значения. Ю.Д.Апресян развил эту идею, подчеркнув важность выявления семантического инварианта лексемы (интегральных сем), который объединяет все её употребления, при этом дифференциальные семы отвечают за конкретные смысловые вариации. Наши данные о полисемии высокочастотных слов подтверждают существование такого инварианта, воспринимаемого носителем языка как общее смысловое ядро.

Концепция Н.Хомского о врождённых концептуальных структурах (пространство, время, причинность и т.д.) объясняет удивительно быстрое усвоение базовой лексики детьми: «ярлычками» становятся слова, а концептуальный аппарат уже заложен биологически. Этот подход согласуется с универсальным характером ядерной лексики в разных языках, поскольку у всех носителей изначально имеется единый набор базовых категорий. Различия же затрагивают преимущественно периферийную лексику, обусловленную культурно-историческими факторами и спецификой языкового употребления.

В контексте нуклеологии слова Л.Блумфилда о лексиконе как «приложении к грамматике» приобретают иное звучание: регулярные, системно значимые элементы (грамматические категории и ключевые лексемы) образуют ядро, тогда как периферийная лексика (редкие, стилистически маркированные или заимствованные слова) менее предсказуема и требует индивидуального описания. Наше сопоставление русского и английского подтверждает эту мысль: в ядре сосредоточены грамматически важные или универсальные понятия, обычно сохранившие исконную природу и устойчивые к изменениям; на периферии больше вариаций и нестабильности.

Результаты созвучны и работам современных корпусных лингвистов (О.Н.Ляшевская, С.А.Шаров, Г.Лич и др.), количественно выявляющих активный словарный запас. В русском, согласно НКРЯ, высокая доля глагольных форм среди наиболее частотных единиц обусловлена многообразием словоизменения; в английском топ-100 преобладают неизменяемые служебные слова. Такие различия важны и для теории (моделирование лексикона), и для практики (составление двуязычных частотных словарей, учебных материалов).

Заключение. Проведённый компаративный анализ лексической нуклеологии русского и английского языков подтвердил существование сравнительно небольшого класса *ядерных лексем*, удовлетворяющих основным коммуникативным потребностям и характеризующихся высокой частотностью. Ядро объединяет универсальные

семантически важные слова, связанные с базовыми категориями человеческого опыта, и языково-специфичные служебные единицы, отражающие грамматический строй. Сопоставление показало, что несмотря на типологические различия (синтетизм и аналитизм, степень словоизменения, наличие/отсутствие артиклей), в обоих языках формируется сходное семантическое ядро, тогда как формальное воплощение (распределение частей речи, полисемия, функциональный спектр) варьируется. Теоретическая значимость результатов состоит в расширении представлений о *лексическом ядре* и механизмах его функционирования. Понятие «лексическая нуклеология» подчёркивает автономность исследования ядровой лексики, которая находится на стыке семантики, статистической лексикографии и типологии. Выявленные факты подтверждают ключевые идеи Виноградова, Апресяна, Лайонза, Хомского, Блумфилда и др., касающиеся семантического инварианта, лексических универсалий и сопряжённости лексикона с грамматикой. Практическая ценность видится в возможностях применения этих данных в лексикографической практике (составление двуязычных словарей базовой лексики), методике преподавания (акцент на наиболее частотных словах) и компьютерной лингвистике (оптимизация алгоритмов перевода и распознавания речи).

Список использованной литературы

1. Bloomfield L. Language. New Yorks Holt, Rinehart and Winston. – 1933. – 566 p.
2. Chomsky N. Aspects of the Theory of Syntax. – MIT press, 2014. – №. 11.
3. Leech G. et al. Word frequencies in written and spoken English: Based on the British National Corpus. – Routledge, 2014. – 300 p.
4. Lyons J. Semantics. – Cambridge university press, 1977. – Т. 2. – 897 p.
5. Swadesh M. The origin and diversification of language. – Routledge, 2017. – P. 176–189.
6. The British National Corpus, XML Edition / Distributed by Oxford University Computing Services – URL: <http://www.natcorp.ox.ac.uk>
7. Апресян Ю. Д. Синонимические средства языка //Избранные труды.–М.: Языки рус. культуры: Восточная литература. – 1995. – 472 с.
8. Васильев М. Е., Милитарёв А. Ю. Глоттохронология в сравнительно-историческом языкознании. Модели дивергенции языков //Модели дивергенции языков [Glottochronology in Comparative-Historical Linguistics and the Models of Linguistic Divergence]\Аспекты компаративистики. – 2008. – С. 509-536.
9. Виноградов В. В. Основные типы лексических значений слова //Вопросы языкознания. – 1953. – Т. 5. – С. 3-29.
10. Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка на материалах Национального корпуса русского языка. – Общество с ограниченной ответственностью" Издательский центр" Азбуковник", 2009.
11. Национальный корпус русского языка – URL: <http://ruscorpora.ru>