



Published in the Russian Federation
 Oriental Studies (Previous Name: Bulletin of the Kalmyk Institute
 for Humanities of the Russian Academy of Sciences)
 Has been issued as a journal since 2008
 ISSN: 2619-0990; E-ISSN: 2619-1008
 Vol. 18, Is. 2, Pp. 483–498, 2025
 Journal homepage: <https://kigiran.elpub.ru>



УДК / UDC 81`3

DOI: 10.22162/2619-0990-2025-78-2-483-498

Моделирование вариативности языковой картины мира русского и арабского языков на основе параллельных корпусных данных (на примере сказок «Тысяча и одна ночь»)

Джаафар Аль Дауд¹, Елена Борисовна Козеренко^{2, 3}

¹ Российский университет дружбы народов (д. 6, ул. Миклухо-Маклая, 117198 Москва, Российская Федерация)

аспирант, младший научный сотрудник



0009-0000-7991-7740. E-mail: [jaafardawood1993\[at\]gmail.com](mailto:jaafardawood1993[at]gmail.com)

² Российский университет дружбы народов (д. 6, ул. Миклухо-Маклая, 117198 Москва, Российская Федерация)

кандидат филологических наук, доцент

³ Федеральный исследовательский центр «Информатика и управление» РАН (д. 44, корп. 2, ул. Вавилова, 119333 Москва, Российская Федерация)

кандидат филологических наук, старший научный сотрудник



0000-0001-7170-6383. E-mail: [kozerenko\[at\]mail.ru](mailto:kozerenko[at]mail.ru)

© КалмНЦ РАН, 2025

© Аль Дауд Д., Козеренко Е. Б., 2025

Аннотация. *Введение.* Данная статья посвящена исследованию языковой картины мира носителей русского и арабского языков. *Цель исследования* — построить двуязычную русско-арабскую модель языковой картины мира как единой концептосферы и исследовать ее вариативность, учитывая семантические и статистические характеристики ключевых концептов и их связей. Для этого решаются задачи по определению лексико-семантических средств, методов создания и анализа параллельных текстовых корпусов, а также по выбору и обработке культурно-значимых текстов на русском и арабском языках. *Материалы и методы.* Основными материалами исследований являются сказки «Тысяча и одна ночь» в оригинале и их переводы на русский язык, выполненные М. А. Салье, — единственный полный перевод этого памятника арабской культуры, осуществленный на русский язык с языка оригинала. Исследование, представленное в данной статье, опирается на структурно-лингвистические и когнитивные модели, такие как фреймы, лексико-семантические поля, частотно-статистические методы для сравнительного анализа лингвистических явлений в русском и арабском языках, конвейерную обработку естественного языка (с помощью инструментов Intelligent Statistical Verbalizer ISV

и Leferent 0.3.0.). В *результате* работы построена двуязычная русско-арабская модель языковой картины мира как единая концептосфера, проведено исследование ее вариативности с учетом семантических и статистических характеристик ключевых концептов (сущностей) и связей (действий). *Выводы.* В работе показано, что лингвистическая модель мира отражает субъективное восприятие объективной реальности, а вариативность языковой картины мира обусловлена многозначностью и неоднозначностью языковых знаков, где контекст определяет выбор конкретного значения.

Ключевые слова: языковая картина мира, концептосфера, русский язык, арабский язык, параллельные корпуса, культурно-значимые тексты, фреймы, лингвистические модели, частотные словари

Благодарность. Выражаем благодарность Российскому университету дружбы народов за поддержку в реализации научных инициатив и содействие активности студентов в публикационной деятельности.

Для цитирования: Аль Дауд Д., Козеренко Е. Б. Моделирование вариативности языковой картины мира русского и арабского языков на основе параллельных корпусных данных (на примере сказок «Тысяча и одна ночь») // Oriental Studies. 2025. Т. 18. № 2. С. 483–498. DOI: 10.22162/2619-0990-2025-78-2-483-498

Parallel Corpus Data in Variability Modeling of Russian and Arabic Linguistic Worldviews: Analyzing Texts of One Thousand and One Nights

Jaafar Al Dawood¹, Elena B. Kozerenko^{2, 3}

¹ RUDN University (6, Miklouho-Maclay St., 117198 Moscow, Russian Federation)

Postgraduate Student, Junior Research Associate



0009-0000-7991-7740. E-mail: jaafardawood1993[at]gmail.com

² RUDN University (6, Miklouho-Maclay St., 117198 Moscow, Russian Federation)

Cand. Sc. (Philology), Associate Professor

³ Federal Research Center 'Computer Science and Control' of the RAS (Bldg. 2, 44, Vavilov St., 119333 Moscow, Russian Federation)

Cand. Sc. (Philology), Senior Research Associate



0000-0001-7170-6383. E-mail: kozerenko[at]mail.ru

© KalmSC RAS, 2025

© Al Dawood J., Kozerenko E. B., 2025

Abstract. Introduction. The paper examines linguistic pictures of the world (LPW) characteristic of native Russian and Arabic speakers. **Goals.** The study attempts a bilingual Russian-Arabic LPW model as an integral conceptual framework — to explore its variability, particular attention be paid to semantic and statistical properties of key concepts and their connections. To facilitate this, the work shall define lexical and semantic means, methods to create and analyze parallel text corpora, as well as ones to select and process culture-significant texts in Russian and Arabic. **Materials and methods.** The work focuses on One Thousand and One Nights (The Arabian Nights) — the original text and its Russian translation by M. Salje who is the one and only to have authored a complete Arabic-to-Russian translation of the literary monument. The study employs structural linguistic and cognitive models, such as frames, lexical-semantic fields. It also addresses statistical methods for comparative analysis of linguistic phenomena in Russian and Arabic, involves pipeline processing of natural language (with the aid of Intelligent Statistical Verbalizer ISV and Leferent 0.3.0.). **Results.** The efforts have yielded a bilingual Russian-Arabic LPW model as an integral conceptual framework,

its variability duly investigated with consideration to semantic and statistical characteristics of key concepts (entities) and connections (actions). *Conclusions.* As is shown, the linguistic model of the world reflects subjective perceptions of objective reality, and the variability of the linguistic picture of the world is rooted in polysemy and ambiguity of linguistic signs where it is the context that specifies the meaning.

Keywords: linguistic worldview, conceptual framework, Russian, Arabic, parallel corpora, culture-significant texts, frames, linguistic models, frequency dictionaries

Acknowledgements. The authors express gratitude to RUDN University for support of scientific initiatives and promotion of student publication activity.

For citation: Al Dawood J., Kozerenko E. B. Parallel Corpus Data in Variability Modeling of Russian and Arabic Linguistic Worldviews: Analyzing Texts of One Thousand and One Nights. *Oriental Studies*. 2025; 18(2): 483–498. (In Russ.). DOI: 10.22162/2619-0990-2025-78-2-483-498



1. Введение

Данная статья посвящена исследованию языковой картины мира (далее — ЯКМ) носителей русского и арабского языков, моделированию вариативности языковой картины мира на основе параллельных лингвистических корпусов текстов и построению *концептосферы* (или ментального пространства исследуемых текстов). Исследование ментальности этнической группы и ее отражение в речевой деятельности носителей рассматриваемого языка является на сегодняшний день одной из самых актуальных проблем в области гуманитарных наук, включая лингвистику и множество филологических дисциплин.

Актуальность данного исследования вытекает из необходимости сопоставительного исследования способов языкового выражения референтной ситуации носителями русского и арабского языков, выявления сходства и различия этих способов и систематического описания языковых средств, наиболее адекватно выражающих ключевые понятия языковой картины мира, используемые носителями этих языков для выражения многоплановой картины мира как проекции ментального пространства. Для русско-арабской языковой пары недостаточно представлены параллельные текстовые корпуса. В результате наших исследований был построен лингвистический ресурс, включающий в себя параллельные тексты, частотные словари и единую концептосферу пространства текстов всех сказок «Тысячи и одной ночи».

Исследуя разнообразные языки, один из основателей современного языкознания В. фон Гумбольдт в своих работах предложил идею, что каждый язык создает у его носителей уникальное видение мира, формируя для своего народа уникальное восприятие реальности [Гумбольдт 1985: 370].

По мнению В. И. Постоваловой, «картина мира» (далее — КМ) представляет собой внутреннее восприятие, а «модель мира» — внешний образ. Модель мира формируется через анализ КМ и отражает осознанное восприятие, тогда как КМ может быть нечеткой и неосознанной [Серебренников и др. 1988: 23]. КМ не просто отображает, но и интерпретирует реальность. Метафора «картина мира» показывает, что как художник создает свою версию реальности, так и человек представляет реальность через язык [Санцевич 2003: 36]. А. Глаз предполагает, что концепция языковой картины мира имеет много общего с лингвистической относительностью [Glaz 2022: 14].

Лингвистические модели исследуют такие вопросы, как определяющие атрибуты концепций, как концепции связаны друг с другом и какие отношения между концепциями являются необходимыми, допустимыми или обязательными [Mansourov, Campara 2011: 233]. А. Эйнштейн отмечал, что человек стремится к понятной и ясной картине мира [Эйнштейн 1967: 153]. В свою очередь О. А. Корнилов определяет КМ как упрощенную модель реального мира, создаваемую человеком [Корнилов 2000: 14].

Предметом данного исследования является языковая картина мира (*концептосфера*), представленная в культурно-значимых текстах как отражение этнической ментальности, а также ее вариативность, т. е. различные способы, которыми языки отображают и интерпретируют реальность, при этом в фокусе внимания данного исследования находятся лексические средства исследуемых языков, их семантические и статистические характеристики.

В процессе исследования была подтверждена *гипотеза* о том, что анализ параллельных лингвистических корпусов позволяет выявить общие и уникальные черты языковой картины мира между различными языковыми системами. Анализ параллельных текстов на русском и арабском языках позволил понять, какие лингвистические феномены и структуры имеются в каждом из языков, а также как эти феномены взаимодействуют и пересекаются в процессе межкультурной коммуникации.

В данной работе мы используем термины «ментальное пространство», «концептуальное пространство», «концептосфера» в значении концептуальной (ментальной) модели как отражения сознания носителей некоторого языка и соответствующей языковой картины мира. Термин «концептосфера» имеет конкретного автора — Д. С. Лихачева: «*В совокупности потенции, открываемые в словарном запасе отдельного человека, как и всего языка в целом, мы можем назвать концептосферами*» [Лихачев 1997: 282].

Языковая картина мира носителей некоторого естественного языка формируется на основе текстов, спонтанно порождаемых носителями языка, а также культурно-значимых текстов. В данной работе мы используем термин «культурно-значимые тексты», под которым понимаем авторские или фольклорные тексты, которые, с одной стороны, отражают существенные черты языковой картины мира носителей языка, а, с другой стороны, сами выступают активными факторами, формирующими эту картину, особенно ее лексическую составляющую, поскольку концепты, фразы, устойчивые выражения из культурно-значимых текстов

становятся частью порождающих механизмов языковой системы.

2. Материалы и методы исследования

Выбор культурно значимых арабских текстов включает в себя рассмотрение нескольких факторов для обеспечения всестороннего представления и актуальности. «Тысяча и одна ночь» выделяется как грандиозное литературное достижение, глубоко влияющее на письменность на протяжении различных исторических эпох. Оно служит культурным мостом между Востоком и Западом и имеет огромное влияние на различные литературные жанры [Jweid 2020: 11].

Для получения лексико-частотных характеристик текстов в наших исследованиях мы используем систему Intelligent Statistical Verbalizer (ISV). Эта система анализа текста выполняет три основные функции: (1) лексический анализ для определения категориальных характеристик лексем; (2) морфологический анализ, позволяющий получить каноническую форму каждого слова (лемму); и (3) подсчет частоты встречаемости слов после получения их канонической формы [АТСИОНД 2020: 113–120]¹.

В анализе текста определенные слова, такие как предлоги, местоимения, союзы и другие лексические единицы, часто называемые «стоп-словами», как правило, являются наиболее частотными и появляются в верхних строках файла встречаемости. Эти слова, несмотря на их частотность, не имеют существенного лексико-семантического значения и могут быть не важны для конкретного исследования, проводимого в наших целях. Таким образом, один из подходов в анализе текста — это отделение этих стоп-слов в специальную категорию при

¹ Простой подсчет количества словоупотреблений может ввести в заблуждение. Это происходит потому, что одно и то же слово может иметь множество различных форм (спряжений по лицам и числам) таких, например, как *читаю*, *читал*, *читать* и т. д. Более правильным подходом является объединение всех этих вариаций под одной «базовой» формой, называемой леммой. Этот процесс нахождения леммы для каждого слова называется лемматизацией [Toporkov, Agerri 2024: 157]. Используя леммы, мы можем получить более четкое представление о том, как часто слова действительно встречаются в тексте.

анализе текстов или полное их удаление вообще [Большаков 2024: 46].

Для моделирования ментального пространства носителей языковой картины мира мы обращаемся к методам и технологиям компьютерной обработки естественного языка. В нашем случае целью моделирования является *ментальное пространство* текста, или *концептосфера*, *концептуальная структура*. Такая концептуальная структура предполагает выявление ключевых сущностей и связей (действий) и установление отношений между ними.

В исследованиях текстов сказок «Тысяча и одна ночь» была применена новая версия семантически-ориентированного процессора (СОЛП) [АТСИОНД 2020: 113–120] — Стенд Leferent 0.3.0. Методы конвейерной обработки естественного языка, включающей морфологический, синтаксический, семантический анализ, организованные в структурированный технологический каркас, способствуют более эффективному анализу текстов, помогая в задачах, таких как разработка онтологий и более глубокое понимание лингвистических данных, построение семантического представления текста как концептосферы. С помощью лингвистического процессора Стенд Leferent 0.3.0 мы получаем также структурированное концептуальное пространство текста сказок «Тысяча и одна ночь» в виде XML-кода.

При исследовании параллельных текстов на русском и арабском языках точкой входа в концептуальное пространство текста для нас являются тексты на русском языке. Им приводятся в соответствие тексты на арабском языке, вначале выравниваемые по предложениям, затем по ключевым и опорным словам, выделяемым сущностям и связям. Стенд Leferent 0.3.0 пока работает для русского и английского языков. Разработка модуля для арабского языка находится в ближайшей перспективе. Поэтому эксперименты в данной работе проведены в смешанном режиме — автоматизированном для русского языка и в ручном для выравнивания с арабскими соответствиями.

Для моделирования и сопоставления семантической структуры лексем, выражающих ключевые концепты языковой картины

мира, используется фреймовый подход. Полученные концептуальные представления текстов в виде опорных концептов и ключевых слов, ключевых действий и связей, частотных словарей, а также структурированных представлений пространства текстов служат материальной основой для создания двуязычных русско-арабских лингвистических ресурсов, предназначенных для многофункционального применения в исследованиях, создания программного обеспечения для систем обработки естественного языка, разработки учебных материалов по различным лингвистическим специальностям, переводу и переводоведению. Двуязычный лингвистический ресурс параллельных культурно-значимых текстов и структурно-семантических описаний ключевых концептов и опорных слов русского и арабского языков реализует *концептосферу* как языковую модель мира.

Полученные концептуальные представления текстов в виде опорных концептов и ключевых слов, ключевых действий и связей, частотных словарей, а также структурированных представлений пространства текстов служат материальной основой для создания двуязычных русско-арабских лингвистических ресурсов, предназначенных для многофункционального применения в исследованиях, создания программного обеспечения для систем обработки естественного языка, разработки учебных материалов по различным лингвистическим специальностям, переводу и переводоведению. Подобный ресурс создается впервые для русско-арабской языковой пары, что обуславливает высокую степень новизны данного исследования.

3. Структурно-семантические модели некоторых ключевых концептов языковой картины мира для русско-арабской языковой пары

А. Вежицкая утверждает, что в русской культуре есть некоторые ключевые концепты, которые занимают особенно важное место [Вежицкая 1999: 282]. Она рассматривает ключевые концепты через призму лексики, подчеркивая, что лексика является чувствительным индикатором культуры

народа [Вежбицкая 2001: 61]. Основываясь на методе выделения «ключевых слов», А. Вежбицкая выделяет три фундаментальных концепта: «судьба», «душа» и «тоска» [Вежбицкая 2001: 36]. На лингвоспецифичность этих концептов для русской языковой картины мира указывает А. А. Зализняк [Зализняк 2006: 207].

Для сопоставительного исследования семантической структуры некоторых ключевых макроконцептов: «судьба», «душа», «тоска», — выделяемых различными авто-

рами для русской языковой картины мира, мы обратились к их аналогам в арабском языке.

В таблице 1 представлена семантическая структура концепта «судьба» — *قدر* [qadar] на арабском языке. Как можно увидеть из представленных примеров, *قدر* имеет весьма обширное пространство возможных значений, в отличие от русского эквивалента *судьба*, что очень наглядно иллюстрирует явление асимметрии языкового знака.

Таблица 1. Семантическая структура концепта *قدر* [qadar] ('судьба') для арабского языка
[Table 1. Semantic structure of the concept *قدر* [qadar] 'fate' in the Arabic language]

№	قدر	Судьба
1	قدر ثمن سلعة بين مقداره	Определил свою стоимость.
2	أعطاه ما يستحق من عناية أو تعظيم	Он даровал ему заслуженный уход или почитание.
3	دبره وفكر في تسويته	Размышлял и планировал его улаживание.
4	تمكن منه وقوي عليه. قدر على النهوض وحده	Справился с ним и укрепился на нем. смог подняться самостоятельно.
5	ضيقه	Сузил его.
6	قاسه به وجعله على مقداره	Оценил его по его заслугам.
7	طبخه في القدر	Приготовил его в кастрюле.
8	قصر. قدر الرجل	Был сокращен. Приговорил мужчину.
9	كفاءة انسان او صفاته الأخلاقية او الأدبية او جدارته، شأن. فنان كبير القدر	Квалификация человека или его моральные, литературные качества или его достоинство, дело. Выдающийся художник.
10	مساوي الشيء من غير زيادة ولا نقصان. هذا قدر ذاك	Это равное, без увеличения или уменьшения. Это равняется тому.
11	مقدار. قدر عمل	Количество. Рабочий объем.
12	ما يعادل ويساوي، بحسب. أسهم بقدر وسائله	То, что равно и соответствует, в зависимости от. Вложил в соответствии с его возможностями.
13	ما يبلغه شيء من أهمية. خطأ بحذر يكلف غالبا	То, что имеет значение. Ошибка в этом может быть дорогой.
14	توزيع موسيقي إلى وحدات زمنية متساوية هي الفواصل. خفف القدر	Распределение музыки на равные временные единицы является перерывом. Снизил громкость.
15	إناء يطبخ فيه الطعام، مرجل، طنجرة، كفت. قدر من نحاس	Емкость, в которой готовится пища, котел, кастрюля, кастрюля, чашка. Медная кастрюля.
16	ما يقدره الله، قضاء يقضي به الله على عباده. ضحية القدر	То, что Бог определяет, решение, которым Бог судит Своих рабов. Жертва судьбы.
17	مبلغ الشيء. بلغ في العلم قدرا	Сумма чего-либо. Достиг в знании определенного уровня.
18	حسب بالتقريب مستعينا بخبرته وبما لديه من معطيات. قدر نفقات	Оценил приблизительно, опираясь на свой опыт и имеющиеся у него данные. Определить расходы.
19	وزن بين احتمالات واختار منها ما يتوقعه. قدر حظوظ النجاح	Взвесил вероятности и выбрал то, что ожидал. Оценил шансы на успех.

20	عين الأهمية والمدى. قدر الأخطار	Определил важность и масштаб. Определил опасности.
21	حكم وفق المقتضى. قدر كتابا حق قدرهم	Судил в соответствии с обстоятельствами. Оценил писателей по достоинству.
22	عرف القيمة الثقافية والأخلاقية أو القدرة على القيام بعمل ما. قدر بنظرة طالب وظيفة	Оценил культурную и моральную ценность или способность выполнить определенную работу. Оценил взглядом кандидата на должность.
23	ظن، افترض. قدر ثروة كبيرة لفلان	Подумал, предположил. Он оценил большое состояние для кого-то.
24	أضمر. قدر في جملة جوابا إيجابيا	Скрыл. Он оценил положительный ответ в предложении.
25	ضيق. قدر على عياله	Сузил. Смог обеспечить свою семью.
26	احترم. قدر التلميذ معلمه	Уважал. Ученик уважительно относился к своему учителю
27	فلان. فكر في تسوية أمره وتدبيره	Такой-то. Размышлял об урегулировании своего дела и его организации.
28	قضى وحكم به	Он вынес решение.

Как мы видим, сфера употребления концепта *قدر* [qadar] 'судьба' в арабском языке весьма многообразна. Проведем сопоставление с системой значений концепта *судьба* в русском языке.

Для русского языка выделяют следующие основные значения:

Судьба — существительное.

1. Стечение обстоятельств: *Судьба столкнула нас с тобой.*

2. Доля, участь: *ничего не знаю о судьбе брата.*

3. Условия дальнейшего существования, будущность: *заботиться о судьбах государства.*

Семантическая структура концепта «душа» для арабского языка приведена в табл. 2. Семантическая структура концепта «душа» для русского языка также представлена в виде табл. 3. Структура значений ключевого концепта русской языковой картины мира «тоска» в сопоставлении с его соответствиями на арабском языке приведена в табл. 4.

Таблица 2. Семантическая структура концепта *روح* 'душа' для арабского языка

[Table 2. Semantic structure of the concept *روح* 'soul' in the Arabic language]

Вещество, извлеченное путем дистилляции и концентрации (дух цветка).	مادة مستخرجة بالتقطير ومركزة (روح الزهر).
Душа и то, что составляет ее жизнь. У вещей нет духа.	ليس للأشياء روح. النفس وما به حياتها.
Призрак, фантом, образ умершего человека.	شبح، طيف، خيال شخص ميت

Таблица 3. Семантическая структура концепта *душа* для русского языка

[Table 3. Semantic structure of the concept *душа* 'soul' in the Russian language]

Душа (имя существительное) Определения	Примеры
Внутренний, психический мир человека, его сознание.	<i>Предан душой и телом своему делу</i>
То или иное свойство характера, а также человек с теми или иными свойствами.	<i>Добрая душа</i>
Вдохновитель чего-н., главное лицо. (ПЕРЕНОСНОЕ ЗНАЧЕНИЕ)	<i>Душа всего дела</i>

О человеке (обычно в устойчивых сочетаниях; разг.).	В доме ни души
В старину: крепостной крестьянин.	Ревизская душа

Таблица 4. Ключевой концепт русской языковой картины мира *Тоска* и соответствия на арабском языке

[Table 4. *Toska* as a key Russian LPW concept and its Arabic equivalents]

Тоска и ностальгия. Желание вернуться в родной дом.	شوق وحنين. توق إلى مسقط الرأس
Сильная тоска. Желание бесконечности.	شوق شديد. توق إلى اللانهاية

Рассмотрим толкования концептов словаре живого великорусского языка» «судьба», «душа», «тоска» в «Толковом В. И. Даля [ТСДО 2008–2024] (табл. 5).

Таблица 5. Концепты *судьба*, *душа*, *тоска* в «Толковом словаре живого великорусского языка» В. И. Даля

[Table 5. Concepts of fate, soul, and longing in The Explanatory Dictionary of the Living Great Russian Language by V. I. Dahl]

Слово	Значение	Контекст / Примечания
Судьба	Суд, судилище, правосудие, расправа.	Означает как судьбу, так и форму высшего правосудия или неизбежный исход.
	Участь, жребий, доля, рок, счастье, предопределение.	Включает понятия предназначения, счастья и роли провидения.
Душа	Бессмертное духовное существо, одаренное разумом и волей.	Олицетворение сущности жизни, совмещающее и духовное, и телесное начало.
	Человек, личность (может относиться к людям обоего пола, либо только к мужскому полу).	Понятие «душа» часто употребляется в смысле «человек», особенно в ревизской и податной документации.
	Совость, духовные качества, внутреннее чувство.	Отражает внутренние, моральные качества человека.
Тоска	Стеснение духа, мучительная грусть, ностальгия, томление.	Глубокая печаль, тоска по родине, иногда приводящая к физическому недомоганию.
	Душевная тревога, беспокойство, боязнь, скука, горе, скорбь.	Выражает чувство внутреннего беспокойства, от лёгкой грусти до сильного горя.

Метод рассмотрения ключевых концептов языковой картины мира в сопоставительном ключе для русско-арабской языковой пары приводит к эффекту выявления дополнительных значений, которые не были бы проявлены в одноязычной ситуации. А. А. Зализняк указывает на этот эффект как на возможный источник многозначности в языке [Зализняк 2006: 206].

Хотя сопоставление языковых картин мира затрудняется различиями в этих картинах, отраженными в языке, однако при конструировании единой концептосферы

для двуязычного текста (оригинала и его перевода) продуктивным оказывается подход, основанный на унификации структур значений для рассматриваемых ключевых концептов и концептосферы текста в целом.

4. Определение семантической структуры и контекстного употребления наиболее частотных «сущностей» и «связей»

Семантические объекты, получившие наименование «сущностей», в естественном языке выражены именами существительными и именными фразами, а глагольная лек-

сика выполняет роль «связей» (между сущностями).

В ходе исследований были выявлены ключевые и наиболее часто встречающиеся концепты из текстов сказок «Тысяча и одна ночь» на русском и арабском. Для русскоязычных текстов был использован метод создания частотного словаря лемм с помощью программного инструмента ISV. Далее

этим концептам на русском языке подбирались эквиваленты в арабских текстах. Частотные словари были сформированы для всех текстов сказок. В таблице 6 приведен фрагмент двуязычного частотного словаря наиболее значимых концептов по текстам сказки о Синдбаде-мореходе. Наиболее частотные глагольные лексемы представлены в табл. 7.

Таблица 6. Наиболее частотные концепты в текстах сказки о Синдбаде-Мореходе
[Table 6. Most frequent concepts in The Tale of Sinbad the Sailor]

№	Частота	Русский язык	Частота	Арабский язык
1	166	остров	166	جزيرة
2	138	человек	138	شخصا
3	132	корабль	132	سفينة
4	116	Синдбад	116	سندباد
5	108	ночь	108	ليلة
6	105	город	105	مدينة
7	100	Аллах	100	الله
8	100	царь	100	ملك
9	71	море	71	البحر
10	68	мореход	68	بحار

Таблица 7. Наиболее частотные «связи» (глагольные лексемы)
в текстах сказки о Синдбаде-мореходе

[Table 7. Most frequent 'connections' (verbal lexemes) in The Tale of Sinbad the Sailor]

№	Частота	Русский язык	Частота	Арабский язык
1	159	быть	159	يكون
2	149	сказать	149	يقول
3	111	стать	111	أصبح
4	90	увидеть	90	انظر
5	59	взять	59	خذ
6	56	знать	56	أعرف
7	54	прийти	54	تأتي
8	52	подняться	52	ارتفاع
9	48	говорить	48	الحديث
10	46	сделать	46	افعل

Таким образом, результаты анализа текста с помощью Интеллектуального статистического вербалайзера (ISV) дают нам частотное представление системы сущностей и связей, образующих ментальное пространство текста, вербализуемое через именные и глагольные лексемы. Проце-

дура концептуального анализа с опорой на отображение сильных позиций лексических единиц в матрице частот словоупотреблений может быть применена в реконструкции ключевых элементов системы смысла при исследовании любого текста. При сопоставлении русского и арабского текстов

на лексико-семантическом уровне мы получаем дополнительное осмысление каждого элемента концептуальной схемы.

5. Двухязычный лингвистический ресурс параллельных культурно-значимых текстов и структурно-семантических описаний ключевых концептов и опорных слов русского и арабского языков

В результате исследований была построена лингвистическая база знаний как новый лингвистический ресурс параллельных культурно-значимых текстов и структурно-семантических описаний ключевых концептов и опорных слов русского и арабского языков, которому было дано имя — АРЛинКор — *أرلينكور* — Арабско-Русский Лингвистический Корпус — *المجموعة* — *اللغوية العربية الروسية*. В данном разделе мы приводим несколько фрагментов АРЛинКор, иллюстрирующих организацию и способы представления лингвистических знаний в этом ресурсе. На рис. 1 приведен фрагмент концептосферы Предисловия к книге сказок «Тысяча и одна ночь» в переводе М. А. Салье и рамочного сюжета — о Шахразаде (*Шехерезаде*) и царе Шахрияре.

Структура текста сказки о Синдбаде-мореходе и ее семантическая структура (*концептосфера*, модель ментального пространства) получены с помощью лингвистического процессора Стенд Leferent 0.3.0.

Сюжет сказки о Синдбаде-мореходе развивается на протяжении 30 ночей, которые были посвящены его семи путешествиям (ночи 536–566). Пространство текста, выстроенное с помощью Leferent 0.3.0, позволяет совершать навигацию по каждой ночи и каждому рассказу об очередном путешествии по гиперссылкам, при этом отображается концептосфера соответствующего раздела сказки.

На следующих рис. 2, 3 представлены фрагменты семантического представления сказки о Синдбаде-мореходе как *концептосферы* в виде графической схемы и XML-кода.

Рассмотрим более подробно ключевые концепты и связи (действия), выделенные лингвистическим процессором, и их соответствия на арабском языке (с транскрипци-

ей), а также *концептосферу*, представленную в виде XML-кода.

Ключевые слова:

ШИРОКАЯ СКАМЕЙКА, СИЛЬНАЯ ЖАРА, ТЯЖЕЛАЯ НОША, ТЯЖЕСТЬ, ЗНОЙ, НОША, ПОВЕЛИТЕЛЬ, ВОРОТ, ВОЗДУХ, СКАМЕЙКА, ХАРУН, ХАЛИФА, АР-РАШИД, ЧЕЛОВЕК, ПЛАТ, СКАЗКА, ГОЛОВА, РЕЧЬ, УТРО, ВРЕМЯ, ДЕНЬ, КУПЕЦ

الكلمات المفتاحية:

مقعد واسع، حرارة شديدة، عبء ثقيل، ثقيل، حرارة، عبء، رب، بوابة، هواء، مقعد، هارون، خليفة، الرشيد، رجل، بلات حكاية، رأس، كلام، صباح، وقت، نهار، تاجر

[alkalimat aldaalatu:

maqead wasie, hararat shadidatun, eib' thaqila, thaqilun, hararatu, eib', raba, bawaabati, hawa'i, maqead, harun, khalifat, alrashida, rajulu, blat, hikayata, rasa, kalami, sabahi, waqat, nahar, tajir]

Ключевые действия:

ПОДМЕСТИ, ВСПОТЕТЬ, ПОЛИТЬ, БЫТЬ, УТОМИТЬСЯ, ПОДЫШАТЬ, НОСИТЬ, НЕСТИ, УСИЛИТЬСЯ, ЗВАТЬ, ПРОХОДИТЬ, ПОЛОЖИТЬ, СТОЯТЬ, ОТДОХНУТЬ, ДОЗВОЛИТЬ, ПРЕКРАТИТЬ, ЗАСТИГНУТЬ, ЖИТЬ, СЛУЧИТЬСЯ

الأفعال الأساسية:

اكتساح، عرق، ماء، كن، متعب، تنفس، احمّل، احمّل، عزز، اتصل، مرر، اللعب، قف، استرح، اسمح، توقف، التقط، حي يحدث

[al'ijra'at alrayiysiatu:

aktisahu, earqa, ma', kun, muteabi, tanafasu, ahmil, ahmil, eazaza, atasala, marra, aleab, qaf, astarha, asmah, tawaquf, altaqatu, hayi, yahduth]

Таким образом представление концептуальной структуры двухязычного пространства текста является синтезом лексико-семантических свойств и контекстного употребления наиболее частотных «сущностей», выраженных именами существительными и именными фразами, и наиболее частотных «связей», представленных глагольной лексикой. Результатом работы семантического лингвистического процессора Стенд Leferent 0.3.0 является графическая, концептуальная, частотная экспликация текста, образ концептосферы как единого ментального пространства текста, выраженного в ключевых и опорных словах и фразах, наиболее частотных лексемах и их структурных связях.

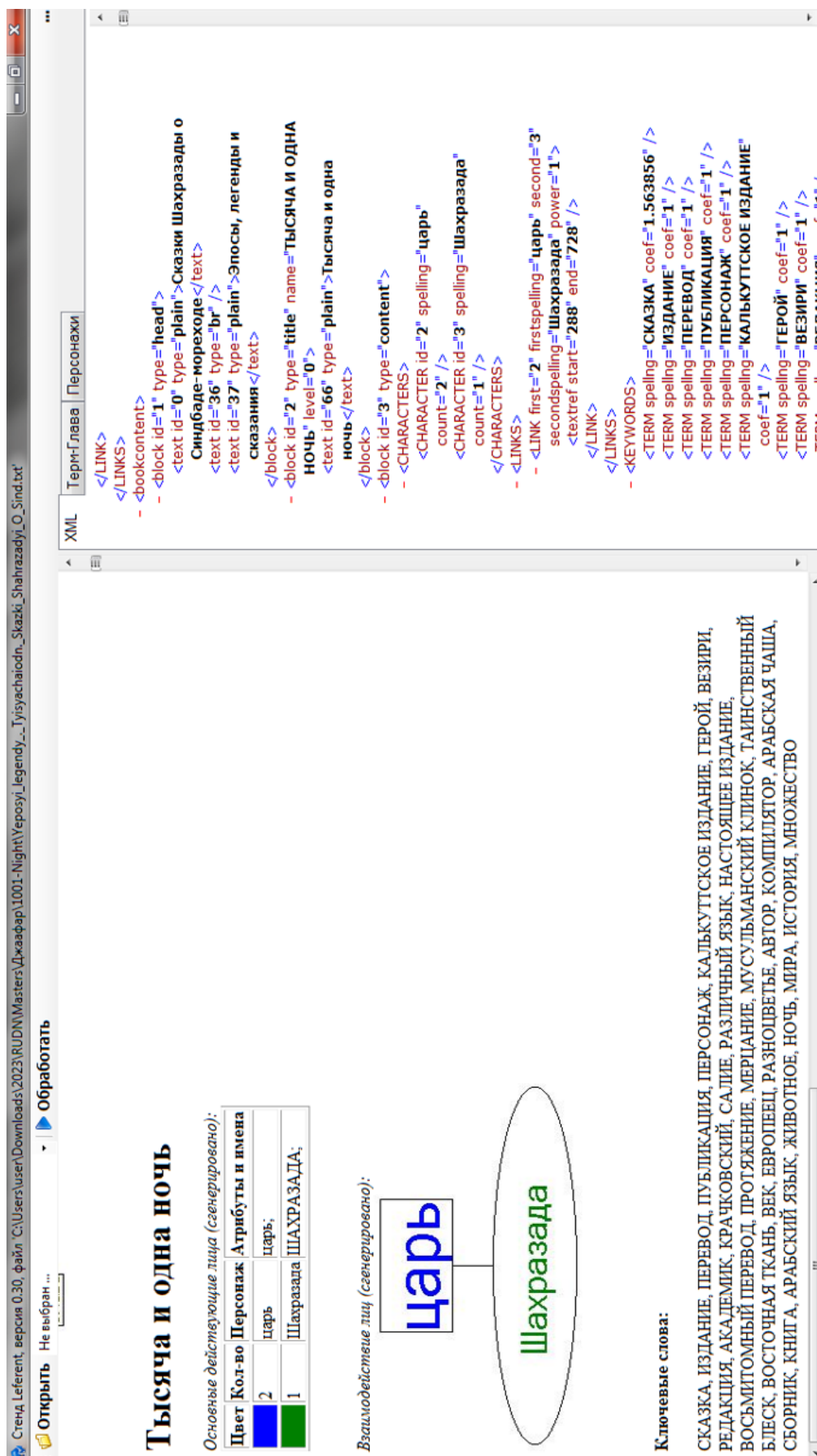


Рис. 1. Фрагмент концептосферы предисловия к книге сказок «Тысяча и одна ночь» и рамочного сюжета о Шахразад и царе Шахрияре [Fig. 1. Fragment of the conceptual framework in the preface to One Thousand and One Nights and the framing narrative about Shahrazad and King Shahriyar]

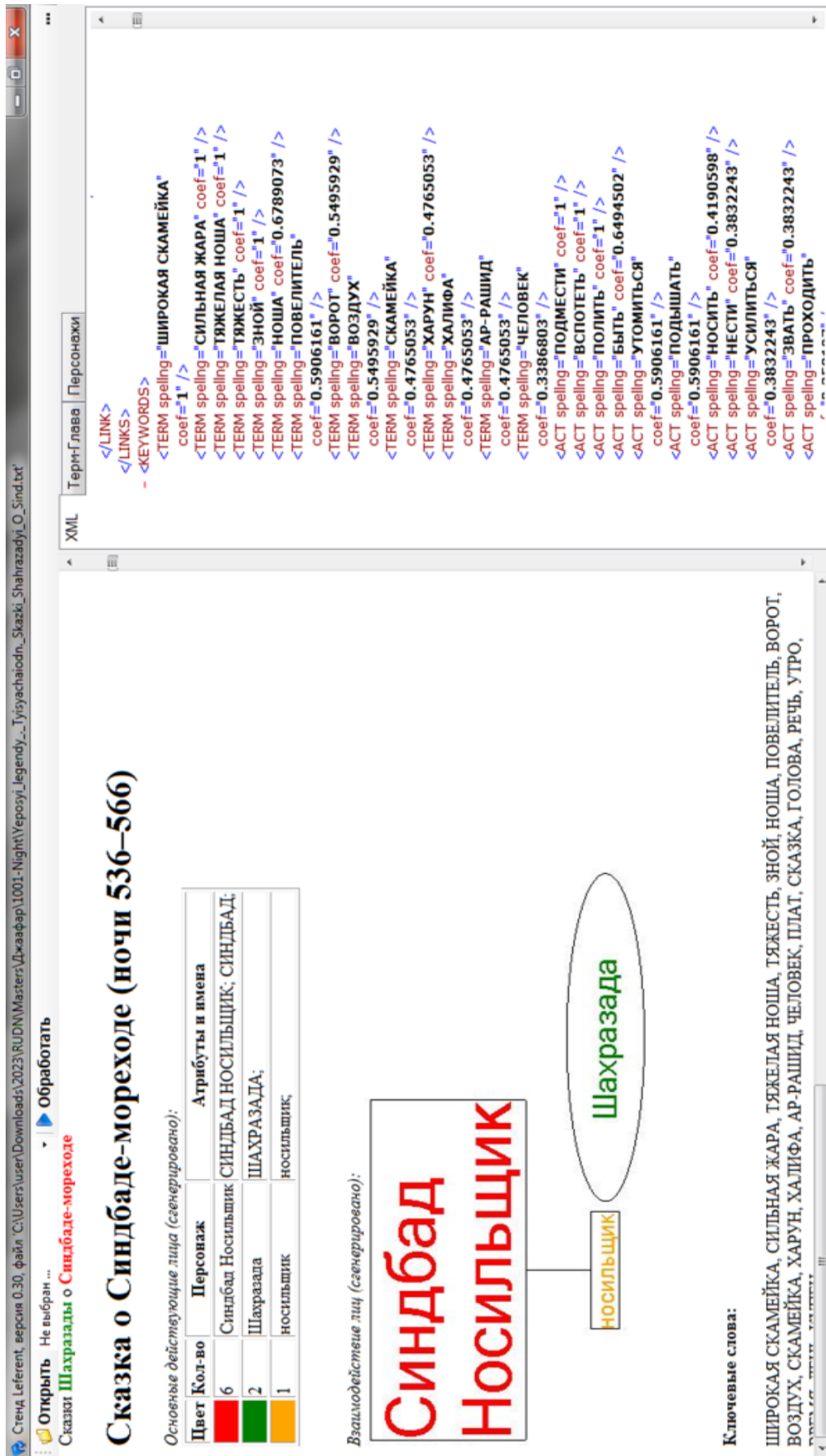


Рис. 2. Семантическое представление фрагмента сказки о Синбаде-мореходе — концептосфера в виде графической схемы и XML-кода [Fig. 2. Semantic representation of a fragment of The Tale of Sinbad the Sailor — conceptual framework in the form of a graphic diagram and an XML Code]

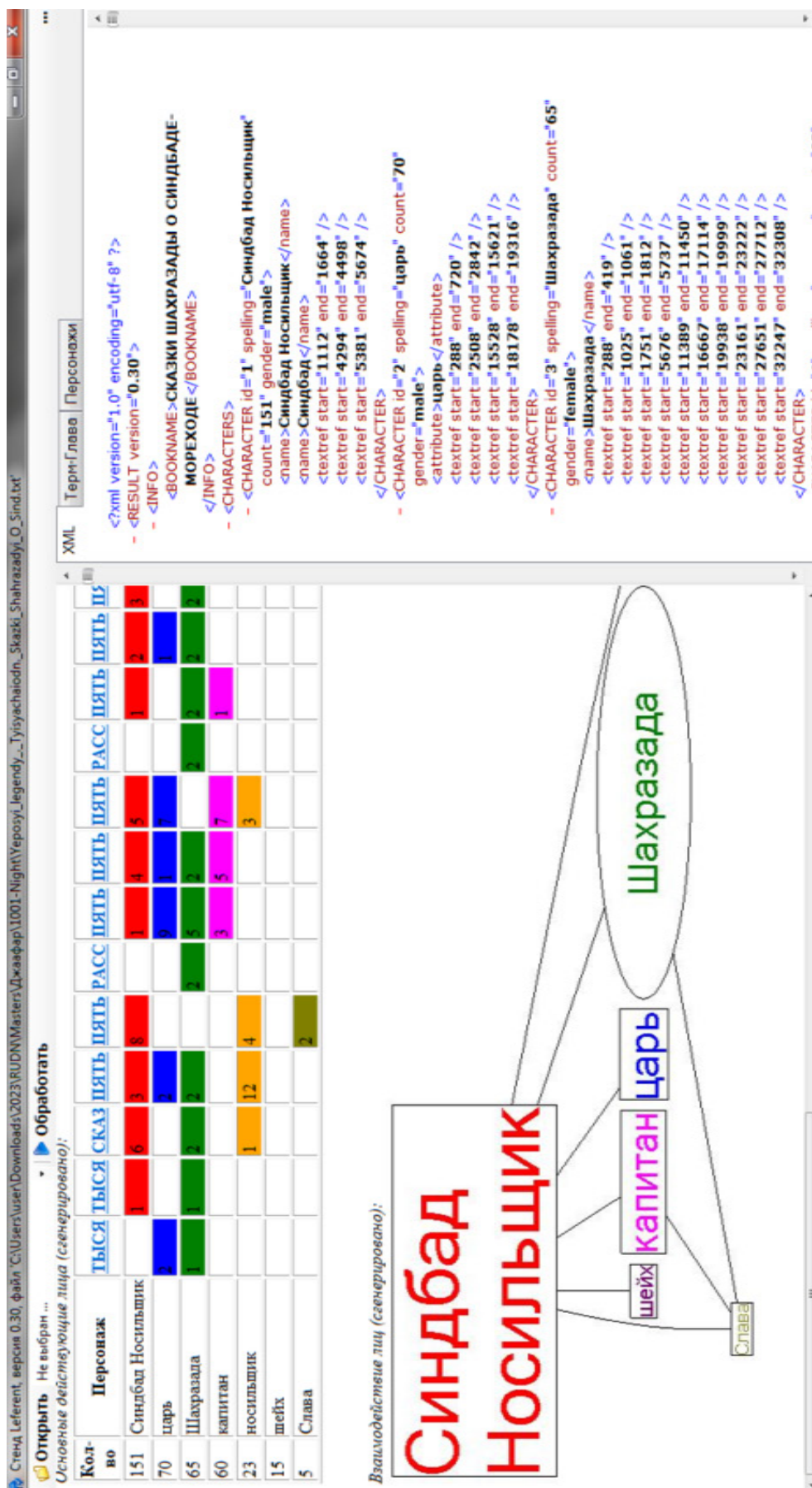


Рис. 3. Семантическое представление фрагмента сказки о Синбаде-мореходе (ночи 536–566) — *концептосфера* в виде графической схемы и XML-кода [Fig. 3. Semantic representation of a fragment of The Tale of Sinbad the Sailor (nights 536–566) — conceptual framework in the form of a graphic diagram and an XML Code]

6. Заключение

Вариативность языковой картины мира обусловлена такими имманентными свойствами естественного языка, как многозначность и неоднозначность языковых знаков. Эти свойства проявляются в устных и письменных текстах, где контекст играет ключевую роль в выборе конкретного значения из всех возможных вариантов языкового знака.

Сопоставление семантической структуры ключевых концептов русского и арабского языков наглядно иллюстрирует явление асимметрии языкового знака и вариативности языковой картины мира.

Культурно-значимые тексты включают авторские или народные произведения, которые, с одной стороны, отражают ключевые аспекты языковой картины мира носителей языка, а с другой стороны, сами оказывают влияние на ее формирование, особенно на лексическую часть. Концепты, выражения и устойчивые фразы из таких текстов становятся частью языковой системы и присутствуют в сознании носителей языка [Mirzoeva, Syurmen 2020: 659]. Таким образом, происходит формирование ментального пространства носителей языка.

В ходе исследований были определены наиболее важные и часто встречающиеся концепты в текстах сказок «Тысяча и одна ночь» на русском и арабском. Для русско-

язычных текстов использовался метод создания частотного словаря лемм с помощью программного инструмента ISV. Затем для выявленных концептов на русском языке подбирались их эквиваленты в арабских текстах. Такие частотные словари были составлены для всех текстов сказок.

Построен двуязычный лингвистический ресурс параллельных культурно-значимых текстов и структурно-семантических описаний ключевых концептов и опорных слов русского и арабского языков. Этот ресурс является новым, ему присвоено название АР-ЛинКор — ارلينكور (англ. ARLinCor) — Арабско-русский лингвистический корпус — الروسية العربية اللغوية المجموعة أرلينكور.

Практическая значимость результатов исследования заключается в том, что разработанные методы, модели и лингвистический ресурс будут востребованной основой для создания и отладки систем машинного перевода, создания *датасетов* (специальных наборов данных) для обучения нейронных сетей и других систем с модулем машинного обучения, для улучшения точности и качества переводов. Полученные результаты могут использоваться в обучении языкам, помогая студентам лучше понять языковую картину мира, ее вариативность и культурные особенности носителей русского и арабского языков.

Литература

- АТСИОНД 2020 — Козеренко Е. Б., Михеев М. Ю., Сомин Н. В. [и др.]. Аналитическая текстология в системах интеллектуальной обработки неструктурированных данных / // Информатика и ее применения. 2020. Т. 14. № 1. С. 113–120. DOI: 10.14357/19922264200115
- Большаков 2024 — Большаков С. С. Анализ записей пользователей социальных сетей для выявления актуальных проблем // Интерактивная наука. 2024. № 1(87). С. 45–50. DOI 10.21661/r-561564.
- Вежбицкая 1999 — Вежбицкая А. Семантические универсалии и описание языков. М.: Языки русской культуры, 1999. 345 с.
- Вежбицкая 2001 — Вежбицкая А. Понимание культур через посредство ключевых слов. М.: Языки славянской культуры, 2001. 287 с.

References

- Kozerenko E. B., Mikheev M. Yu., Somin N. V. et al. Analytical textology in intelligent processing systems for unstructured data. *Informatics and Applications*. 2020. Vol. 14. No. 1. Pp. 113–120. (In Russ.) DOI: 10.14357/19922264200115
- Bolshakov S. S. Social network users' posts' analysis for actual problems' detection. *Interactive Science*. 2024. No. 1(87). Pp. 45–50. (In Russ.) DOI 10.21661/r-561564.
- Wierzbicka A. Semantic Universals and Language Description. Moscow: Yazyki Russkoy Kultury, 1999. 345 p. (In Russ.)
- Wierzbicka A. Understanding Cultures through Their Key Words. Moscow: Yazyki Slavyanskoy Kultury, 2001. 287 p. (In Russ.)

- Гумбольдт 1985 — *Гумбольдт В. Ф.* Язык и философия культуры. М.: Прогресс, 1985. 452 с.
- Корнилов 2000 — *Корнилов О. А.* Языковые картины мира как отражения национальных менталитетов: дисс. ... д-ра культурологии. М., 2000. 460 с.
- Зализняк 2006 — *Зализняк А. А.* Многозначность в языке и способы ее представления. М.: Языки славянских культур, 2006. 672 с.
- Лихачев 1997 — *Лихачев Д. С.* Концептосфера русского языка // *Русская словесность: антология* / под ред. В. П. Нерознака. М.: Academia, 1997. С. 280–287.
- Санцевич 2003 — *Санцевич Н. А.* Моделирование вариативности языковой картины мира на основе двуязычного корпуса публицистических текстов (Метафоры и семантические оппозиции): дисс. ... канд. филол. наук. М., 2003. 268 с.
- Серебренников и др. 1988 — Роль человеческого фактора в языке: Язык и картина мира / Б. А. Серебренников, Е. С. Кубрякова, В. И. Поставалова и др. М.: Наука, 1988. 216 с.
- ТСДО 2008–2024 — Толковый словарь Даля онлайн [электронный ресурс] // URL: <https://www.slovardalja.net/word.php?wordid=7657> (дата обращения: 24.09.2024).
- Эйнштейн 1967 — *Эйнштейн А.* Собрание научных трудов. В 4 тт. / под ред. И. Е. Тамма, Я. А. Смородинского, Б. Г. Кузнецова. М.: Наука, 1965. 630 с.
- Glaz 2022 — *Glaz A.* Linguistic worldview(s): approaches and applications. New York London: Routledge, Taylor & Francis Group, 2022. 240 с.
- Jweid 2020 — *Jweid A. N. A.* The reception of The Arabian Nights in world literature // *Studies in Literature and Language*. 2020. Vol. 22. № 1. Pp. 10–15.
- Mansourov, Campara 2011 — *Mansourov N., Campara D.* System assurance: beyond detecting vulnerabilities. Campara: Elsevier, 2011. 346 с.
- Mirzoeva, Syurmen 2020 — *Mirzoeva L., Syurmen O.* Precedent Text as a Special Kind of Code in the Internet Communication // *Communication Trends in the Post-Literacy*
- Humboldt W. von. Language and Philosophy of Culture. Moscow: Progress, 1985. 452 p. (In Russ.)
- Kornilov O. A. Linguistic Worldview and Manifested National Mentality. Dr. Sc. (culturology) thesis. Moscow, 2000. 460 p. (In Russ.)
- Zaliznyak A. A. Polysemy and Its Representation Means. Moscow: Yazyki Slavyanskikh Kultur, 2006. 672 p. (In Russ.)
- Likhachev D. S. Conceptosphere of the Russian language. In: Neroznak V. P. (ed.) The Russian Word: An Anthology. Moscow: Academia, 1997. Pp. 280–287. (In Russ.)
- Santsevich N. A. Using a Bilingual Journalistic Corpus for Variability Modeling of a Linguistic Worldview: Metaphors and Semantic Oppositions. Cand. Sc. (philology) thesis. Moscow, 2003. 268 p. (In Russ.)
- Serebrennikov B. A., Kubryakova E. S., Postovalova V. I. et al. Human Factor in Language: Language and Worldview. Moscow: Nauka, 1988. 216 p. (In Russ.)
- Душа [Soul]. On: [Dal V. I.] ONLINE Explanatory Dictionary of the Living Great Russian Language. Available at: <https://www.slovardalja.net/word.php?wordid=7657> (accessed: 24 September 2024). (In Russ.)
- Einstein A. Collected Scientific Writings. In 4 vols. I. Tamm, Ya. Smorodinsky, B. Kuznetsov (eds.). Moscow: Nauka, 1965. 630 p. (In Russ.)
- Glaz A. Linguistic Worldview(s): Approaches and Applications. New York, London: Routledge, Taylor & Francis Group, 2022. 240 p. (In Eng.)
- Jweid A. N. A. The reception of The Arabian Nights in world literature. *Studies in Literature and Language*. 2020. Vol. 22. No. 1. Pp. 10–15. (In Eng.) DOI: 10.3968/12025
- Mansourov N., Campara D. System Assurance: Beyond Detecting Vulnerabilities. Amsterdam: Elsevier, 2011. 346 p. (In Russ.)
- Mirzoeva L., Syurmen O. Precedent text as a special kind of code in the Internet communication. In: Simpson R. et al. (eds.) *Communication Trends in the Post-Literacy Era*:

Era: Polylingualism, Multimodality and Multiculturalism As Preconditions for New Creativity. Ekaterinburg: Ural University Press, 2020. C. 648–660.

Toporkov, Agerri 2024 — *Toporkov O., Agerri R.* On the Role of Morphological Information for Contextual Lemmatization // *Computational Linguistics*. 2024. T. 50. № 1. C. 157–191.

Polylingualism, Multimodality and Multiculturalism as Preconditions for New Creativity. Yekaterinburg: Ural University Press, 2020. Pp. 648–660. (In Eng.)

Toporkov O., Agerri R. On the role of morphological information for contextual lemmatization. *Computational Linguistics*. 2024. Vol. 50. No. 1. Pp. 157–191. (In Eng.)

