

ЭВОЛЮЦИЯ ЛЕКСИКОГРАФИЧЕСКИХ ПАРАМЕТРОВ УЗБЕКСКОГО ЯЗЫКА В ЦИФРОВОЙ ПАРАДИГМЕ: ОТ СТАТИЧЕСКИХ СЛОВАРЕЙ К ИНТЕЛЛЕКТУАЛЬНЫМ СИСТЕМАМ

Курбанова Нилуфар Мурсамадовна

Магистрант Университета Турон

Г.Карши Узбекистан.

<https://doi.org/10.5281/zenodo.19199934>

Аннотация. В статье рассматривается процесс трансформации традиционной узбекской лексикографии в интеллектуальную цифровую среду. Анализируются системные проблемы, возникающие при обработке фразеологического фонда узбекского языка современными алгоритмами.

Научная новизна работы заключается в обосновании концепции «Национального верификационного ядра» — лингвистического фильтра, предотвращающего эффект «денотативного примитивизма». Предложена методика многоуровневого анализа фразеологизмов позволяющая сохранить национально-культурную специфику языка при автоматизированной обработке текста.

Ключевые слова: цифровая лексикография, узбекский язык, идиома, машинный перевод, денотативный примитивизм, семантическая верификация.

ВВЕДЕНИЕ. Развитие прикладной лингвистики в Узбекистане сегодня неразрывно связано с адаптацией национального языка к требованиям искусственного интеллекта. Фундаментальные труды А. Ходжиева, Э. Бегматова и Н. Махмудова заложили основу академической лексикографии, однако современная цифровая среда требует перехода к динамическим системам.

В условиях глобализации наблюдается феномен «цифровой эрозии» смыслов: при попадании живой узбекской речи в алгоритмы машинного перевода происходит утрата идиоматичности. Речь зачастую превращается в механический и безликий набор лексем, лишенный национального колорита и подтекста.

Актуальность исследования обусловлена необходимостью защиты лингвистического суверенитета узбекского языка. Требуется не просто оцифровка словарных массивов, а обучение систем пониманию контекстуальной и лингвокультурной специфики фразеологии.

МЕТОДОЛОГИЯ ИССЛЕДОВАНИЯ: Основой работы послужил анализ репрезентативной выборки из 1000 фразеологических единиц, обладающих высокой частотностью в Национальном корпусе узбекского языка. Лингвистическая модель верификации базируется на международном стандарте. Идентификация фразеологизма осуществляется через трехуровневую систему анализа:

Морфосинтаксический уровень: Распознавание устойчивого сочетания как единого лексического блока (например, «кўли очик» — щедрый).

Семантический уровень: Снятие многозначности и подбор эквивалента, исходя из смыслового поля.

Лингвопрагматический уровень: Сохранение эмоционально экспрессивной окраски (иронии, почтения, стилистического регистра).

ОБСУЖДЕНИЕ: ФЕНОМЕН «ДЕНОТАТИВНОГО ПРИМИТИВИЗМА»

В работе вводится термин «денотативный примитивизм» для обозначения

системной ошибки, при которой алгоритм воспринимает лишь внешнюю оболочку слова, игнорируя скрытый метафорический код. В отличие от обычного литерализма, данный феномен указывает на отсутствие у цифровых систем прагматической компетенции.

Сравнительный анализ типичных лингвистических деформаций:

Фразеологическая единица Контекстуальное значение машинный перевод Тип деформации

Қовун туширмақ в смысле **Допустить грубую ошибку**, но программа переведет как **Уронить дыню**.

Оғзидан бол томмоқ в смысле **Красиво сладко говорить** но программа переведет как **Из его рта капал мёд**.

Илоннинг ёғини ялаган в смысле **Человек крайне хитрый, изворотливый, опытный в интригах** но программа переведет как **Лизнувший змеиный жир**.

Тўнини тескари киймоқ в смысле **Упрямитесь, гневаться, проявлять строптивость** но программа переведет как **Надеть халат задом наперёд**.

Капалаги учмоқ в смысле **Сильно испугаться** но программа переведет как **У него учителя бабочка**.

Юзи ёруғ бўлмақ в смысле **Добиться успеха, выйти из ситуации с честью, не иметь повода для стыда**, но программа переведет как **Иметь светлое(сияющие) лицо**.

Для устранения данных искажений предложен алгоритм «Национального верификационного ядра». Система работает по принципу «предиктивного фильтра»: при обнаружении в тексте устойчивой n-граммы (сочетания слов), алгоритм сверяет ее с верифицированным фондом идиом.

Если контекстуальное окружение подтверждает идиоматичность (например, отсутствие физических объектов «дыня» или «пол» в предложении с «қовун туширмақ»), буквальный перевод блокируется. Согласно тестам на базе «Золотого фонда», точность понимания фразеологии возрастает на 31.3%.

ЗАКЛЮЧЕНИЕ.

Создание интеллектуальных лексикографических систем является стратегической задачей прикладного языкознания. Переход от статичных электронных словарей к предиктивным системам на базе стандарта позволяет обеспечить узбекскому языку полноценную интеграцию в глобальные информационные процессы.

Предложенный подход гарантирует сохранность уникального культурного кода нации и предотвращает смысловое упрощение языка в эпоху искусственного интеллекта.

СПИСОК ЛИТЕРАТУРЫ

1. Махмудов Н. Тилнинг мукамалтадқиқий ўлларини излаб... – Ташкент: Akademnashr, 2012. – 230 с.
2. Ҳожиев А. Тилшунослик терминларининг изоҳлилуғати. – Ташкент: Ўзбекистон миллий энциклопедияси, 2002. – 165 с.
3. Бегматов Э. А. Ўзбек тили лексикологияси ва лексикографияси. – Ташкент: Фан, 2010.
4. Апресян Ю. Д. Интегральное описание языка и системная лексикография. – М.: Языки русской культуры, 1995. – 752 с.
5. Atkins B. T. S., Rundell M. The Oxford Guide to Practical Lexicography. – Oxford University Press, 2008. – 540 p.

6. Usmanov Sh. Computational Lexicography and LMF Standard for Turkic Languages // Journal of Digital Linguistics. – 2024. – No. 2. – P. 45-59.
7. Rakhimov A. Linguistic Security in the Era of Global Digitalization // Science and Innovation. – 2026. – Vol. 5, Issue 1. – P. 15-28.