ПРИНЦИПЫ ФОРМИРОВАНИЯ ЭЛЕКТРОННОГО РУССКО-УЗБЕКСКОГО УЧЕБНОГО СЛОВАРЯ НА ОСНОВЕ КОРПУСНЫХ ДАННЫХ

Абдурахматова Мушаррафхон Абдували кизи

Аннотация. В статье предлагается проектная модель электронного русско-узбекского учебного словаря, ориентированного на потребности узбекоязычных изучающих русский язык. Рассматриваются критерии отбора лексики, принципы разграничения значений, представления сочетаемости, грамматических характеристик и корпусных примеров. Обосновывается необходимость сочетания частотных показателей с учебной значимостью, контекстной репрезентативностью и экспертной проверкой переводных соответствий. Описывается структура словарной статьи и предлагается процедура оценки качества словаря. Работа носит теоретико-проектный характер: корпусный эксперимент и пользовательское тестирование представлены как последующие этапы исследования.

Ключевые слова: компьютерная лексикография, учебный словарь, русский язык как иностранный, узбекский язык, корпусная лингвистика, частотность, коллокации, электронный словарь.

ANNOTATION. The article proposes a design framework for an electronic Russian–Uzbek learner's dictionary based on corpus evidence. It discusses headword selection, sense discrimination, collocations, grammatical information, contextual examples, and quality assessment. The framework is conceptual and requires subsequent corpus-based and user evaluation.

Keywords: bilingual lexicography, learner's dictionary, corpus linguistics, Russian as a foreign language, Uzbek, collocations.

Введение

Развитие электронных лексикографических ресурсов изменяет способы представления словарной информации и её использования в обучении иностранным языкам. Для узбекоязычного учащегося русско-узбекский учебный словарь должен не только сообщать перевод слова, но и помогать выбирать значение в контексте, строить нормативные словосочетания и учитывать грамматические особенности русского языка. В обычном двуязычном перечне соответствий сведения об управлении, сочетаемости и стилистических ограничениях часто остаются недостаточно явными. В результате знание переводного эквивалента не гарантирует правильного употребления слова.

Научная проблема состоит в определении того, как корпусные данные могут быть преобразованы в педагогически доступную и лингвистически точную словарную статью. Объект исследования — электронная учебная лексикография русского и узбекского языков. Предмет — критерии отбора и способы представления лексических единиц русского языка в двуязычном словаре для учебных целей. Цель — разработать систему принципов построения такого ресурса с опорой на корпусные свидетельства.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

Для достижения цели решаются задачи: определить параметры отбора заголовочных единиц; описать алгоритм разграничения значений и подбора эквивалентов; установить формат грамматической и сочетаемостной информации; предложить критерии отбора примеров и схему экспертной оценки. Методы исследования — сопоставительный лексико-семантический анализ, лексикографическое моделирование, анализ корпусных контекстов и проектирование процедуры пользовательской проверки.

1. Корпусный подход и учебная направленность словаря

Корпус позволяет наблюдать употребление слова в естественных контекстах, оценивать частотность, выявлять устойчивые сочетания и сопоставлять значения. Однако механическое включение наиболее частотных слов не создаёт полноценного учебного словаря. Высокочастотные служебные единицы, многозначные глаголы и тематически значимая лексика требуют разных способов описания. Следовательно, частотность должна выступать одним из критериев, но не единственным основанием отбора.

Для каждого кандидата в словник предлагается учитывать: (1) частоту и распределение по жанрам; (2) соответствие уровню владения языком; (3) коммуникативную востребованность; (4) грамматическую и словообразовательную значимость; (5) трудность для узбекоязычного учащегося; (6) наличие надёжных контекстов и проверяемых переводных соответствий. Для учебного словаря начального уровня предпочтительны нейтральные единицы повседневного общения, но список должен корректироваться тематикой программы.

Корпусная выборка нуждается в контроле состава: тексты новостей, художественной литературы и устной речи различаются по лексике. Рекомендуется отдельно фиксировать источники и доли жанров, а также не смешивать показатели разных корпусов без нормализации. Частотность следует выражать в употреблениях на миллион слов; при этом нужно учитывать лемматизацию и омонимию.

2. Разграничение значений и переводные эквиваленты

Многозначность представляет особую трудность для двуязычного словаря. Например, русский глагол «идти» в сочетаниях «идти в школу», «идёт дождь» и «идёт урок» реализует различные значения и не должен сопровождаться одним универсальным узбекским переводом. Словарная статья должна делиться на значения, каждое из которых подтверждается типичным контекстом, имеет отдельный эквивалент и грамматическую помету, если она необходима.

Переводные соответствия рекомендуется устанавливать не по изолированной форме, а по значению в конкретной конструкции. В случаях неполной эквивалентности следует давать краткое толкование, а не искусственно подбирать одно слово. Желательно также обозначать ограничения употребления, типичные ошибки переноса и различия в управлении.

3. Представление сочетаемости и грамматических характеристик

Словарь учебного назначения должен показывать, с какими словами и в каких грамматических формах употребляется лексема. Для глагола существенны вид, управление и основные формы; для существительного — род, тип склонения, формы множественного числа и трудные падежные формы; для прилагательного — согласование и распространённые сочетания. Коллокации необходимо отбирать по сочетанию статистической устойчивости и педагогической полезности, а не только по формальному показателю совместной встречаемости.

Особенно полезны сопоставительные комментарии к конструкциям, в которых перевод не передаёт русскую грамматическую модель напрямую. Например, «помогать кому?» требует

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

дательного падежа: «помогать другу». Такой комментарий связывает лексическое значение с практическим построением фразы. При этом не следует объявлять всякое отличие русского и узбекского языков причиной ошибки без анализа реальных ученических текстов.

4. Модель электронной словарной статьи

Предлагается модульная структура, допускающая краткий просмотр для начинающего пользователя и раскрытие подробных сведений по запросу. Ниже приведён иллюстративный макет: он не является результатом автоматического извлечения из корпуса.

Поле	Пример заполнения
Заголовочная единица	помогать
Часть речи и грамматика	глагол, несовершенный вид; помогу? — нет, форма «помогу» относится к совершенному виду «помочь»
Значение	оказывать помощь кому-либо
Узбекский эквивалент	yordam bermoq
Управление	помогать кому? (дательный падеж)
Сочетаемость	помогать другу; помогать родителям; помогать с уроками
Учебный пример	Я помогаю младшему брату делать уроки.
Предупреждение	Не: «помогать брата»; правильно: «помогать брату».

Пример показывает необходимость раздельного хранения леммы, грамматических сведений, значений, переводов, моделей управления и контекстов. Для публикации словаря каждое поле должно пройти лингвистическую проверку. В частности, грамматические сведения нельзя автоматически переносить между видовыми парами глаголов.

5. Отбор корпусных примеров и цифровая реализация

Корпусный пример должен быть достоверным, понятным и репрезентативным для описываемого значения. Рекомендуется исключать контексты с чрезмерно сложной синтаксической структурой, неоправданной культурной зависимостью и потенциально чувствительными данными. Если пример упрощён редактором, его следует помечать как адаптированный, а не представлять дословной корпусной цитатой.

Техническая модель предполагает хранение каждой словарной единицы в структурированной базе с полями идентификатора, леммы, грамматических признаков, значений, эквивалентов, коллокаций, примеров, источника, уровня сложности и даты проверки. Поиск должен поддерживать словоформы, кириллическую и латинскую запись узбекского языка и фильтрацию по учебному уровню. Для воспроизводимости важно фиксировать версию корпуса и правила извлечения данных.

6. Проверка качества и ограничения

Качество словаря предлагается оценивать на двух уровнях. Лингвистическая экспертиза устанавливает корректность значений, переводов, управления, помет и примеров. Пользовательская проверка показывает, помогает ли статья выполнить конкретную задачу: понять текст, выбрать эквивалент или самостоятельно составить предложение. Возможные

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.

показатели — доля верно выполненных заданий, время поиска, число ошибок после обращения к статье и субъективная понятность объяснения.

Для экспериментальной проверки можно сформировать пилотный словник из 100–200 единиц, отобранных по заранее установленным критериям, и сравнить результаты выполнения учебных заданий с использованием разработанного ресурса и обычного двуязычного словаря. Размер выборки и статистический план необходимо обосновать до сбора данных. Поскольку такая проверка пока не проведена, утверждения о превосходстве предлагаемой модели были бы преждевременными.

Заключение

Корпусно ориентированный электронный русско-узбекский учебный словарь следует рассматривать как систему взаимосвязанных сведений о значении, переводе, грамматике и употреблении слова. Предложенные принципы предполагают сочетание частотного анализа, учебной релевантности, контекстного описания значений и экспертной проверки. Практическим результатом настоящей теоретико-проектной работы является модель словарной статьи и набор критериев отбора и оценки. Дальнейший этап исследования — формирование корпусной выборки, создание пилотного словника и проверка его эффективности в реальной учебной ситуации.

Список литературы

1. Atkins B. T. S., Rundell M. The Oxford Guide to Practical Lexicography. Oxford: Oxford University Press, 2008.
2. Kilgarriff A., Rychlý P., Smrž P., Tugwell D. The Sketch Engine // Proceedings of EURALEX 2004. 2004. P. 105–116.
3. Sinclair J. Corpus, Concordance, Collocation. Oxford: Oxford University Press, 1991.
4. Nation I. S. P. Learning Vocabulary in Another Language. 2nd ed. Cambridge: Cambridge University Press, 2013.
5. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating Language Structure and Use. Cambridge: Cambridge University Press, 1998.
6. Tarp S. Lexicography in the Borderland between Knowledge and Non-Knowledge. Tübingen: Max Niemeyer Verlag, 2008.

© Ushbu material Creative Commons Attribution 4.0 International (CC BY 4.0) litsenziyasi asosida tarqatiladi.