

СИНОНИМАЙЗЕР (СИНОНИМИЗАТОР) В ОБРАЗОВАТЕЛЬНОМ КОРПУСЕ УЗБЕКСКОГО ЯЗЫКА

SYNONYMIZER IN THE EDUCATIONAL CORPUS OF THE UZBEK LANGUAGE

Абжалова Манзура Абдурашетовна, кандидат наук по филологии (PhD). Ташкентский государственный университет узбекского языка и литературы имени Алишера Навои, Узбекистан.

Abjalova Manzura Abdurashetovna, Doctor of Philosophy in Philology (PhD). Tashkent State University of Uzbek Language and Literature named after Alisher Navoi, Uzbekistan.

abjalova.manzura@gmail.com

ORCID: 0000-0002-1927-2669



Сегодня изучение и преподавание языков с использованием цифровых технологий является актуальным вопросом. В связи с этим создаются лингводидактические платформы, такие как различные словари, мобильные приложения и компьютерные программы. Образовательный корпус узбекского языка, созданный с учетом потребности в современных технологических средствах языкового образования, решает такие задачи, как повышение эффективности обучения, предоставление удобных дидактических возможностей, содействие развитию словарного запаса при изучении языка. То, что это корпус служит единой образовательной лингво-информационной системой, определяется наличием учебников в его базе данных. В базу данных также входит словарь синонимов, что позволяет корпусу выступать в качестве синонимизатора. В данной статье рассматривается синонимайзер, принцип его работы и начальный процесс работы синонимайзера в узбекском образовательном корпусе.

Ключевые слова: узбекский язык, учебный корпус, синонимайзер, синонимизатор, синонимы.

Today, learning and teaching languages is a global problem. Therefore, linguodidactic platforms are being created, such as various dictionaries, mobile applications and computer programs. The educational corpus of the Uzbek language, created taking into account the need for modern technological means of language education, solves such problems as increasing the effectiveness of learning, providing convenient didactic opportunities, promoting the development of vocabulary in learning the language. The fact that this corpus serves as a unified educational linguistic information system is determined by the presence of textbooks in its database. The database also includes a synonym dictionary, which allows the corpus to act as a synonymizer. This article discusses the synonymizer, the principle of its operation and the initial process of the synonymizer operation in the Uzbek educational building.

Key words: Uzbek language, educational corpus, synonymizer, synonyms.

Введение

Узбекский язык – богатый язык. Возникает вопрос: измеряется ли богатство языка количеством слов или его исторической значимостью? Может быть, есть другой критерий? Конечно, есть. Богатство языка заключается не только в большом количестве слов, но и в разнообразии их значений. Новые смысловые оттенки придают языку гибкость, живость и выразительность. В языке много различных омонимов, синонимов, антонимов, паронимов, которые делают речь образной, красноречивой, экспрессивной, помогают избежать повторения одних и тех же слов, позволяют эмоционально выразить мысль. Каждое слово имеет альтернативное значение в зависимости от функции стиля речи (разговорный / диалектный, художественный, официальный, научный, публицистический). Функциональные стили речи – разновидности литературного языка, которые обусловлены различными сферами общения. Они создаются в зависимости от целей и задач общения и соответственно различаются отбором языковых средств. Знание синонимических богатств родного языка – необходимое условие речевой культуры человека.

Целью статьи является представление принципа работы и процесса работы синонимайзера в Узбекском образовательном корпусе. Материалом статьи послужили материалы Узбекского образовательного корпуса.

Использование синонимайзера в образовательном корпусе узбекского языка

Синонимия в лингвистике является универсальным языковым явлением. Синонимический ряд является связующим звеном, коммуникативным центром между разными речевыми сферами, разными стилями речи (Корсакова 2010: 105-106).

Известно, что существует два основных типа синонимов: лексические синонимы и контекстные синонимы.

Лексические синонимы, в свою очередь, имеют полные (абсолютные) и частичные (семантические, методологические и методолого-семантические) синонимы. В то время как лексические синонимы формируют синонимию в результате близости и согласованности в семантике слов, а контекстуальные синонимы формируются благодаря лексической согласованности, которая возникает только в контексте текста. Хотя типы синонимов классифицируются в источниках по-разному, в целом можно выделить следующие типы:

Схема 1. Типы синонимов.



Контекстные или текстовые синонимы также называются авторскими синонимами²⁸, и они обычно создают синонимию только в контексте определенного контекста и не могут быть синонимами вне контекста. Их главная задача заключается в описании события, а не в его констатации. Поэтому составление словаря текстовых синонимов – сложный процесс²⁹. К типу лексических синонимов относятся также квазисинонимы в узбекском языке, которые еще не изучены глубоко. Квазисинонимы (ложные синонимы, частичные синонимы³⁰ – это слова, которые относительно похожи по значению, но во всех случаях не являются взаимозаменяемыми. Известно, что лексические синонимы – это в основном лексические единицы со схожими значениями, принадлежащие одной и той же группе слов. Квазисинонимы также образуются синонимами слов, принадлежащие разным группам слов. Например: лемматизация (анализ слова до основной морфемы; определение основы) – (N – существительное) ↔ лемматизациялаш – (Vh – Глаголь название действия); регенерация (N – существительное) ↔ қайта тикланиш (Vh – Глаголь название действия); дахшат (N – существительное) ↔ zo‘r (Adj – прилагательное) (в стиле устного общения).

Как правило, синонимы используются в речи для следующих целей, а в некоторых местах даже необходимы:

- в выражении идеи более четко и точно (для сравнения: *чет эл* и *хориж*);
- в более ярком выражении мысли, эмоциональной окраске;
- во избежание тавтологии (повторение - многократное употребление слова в предложении).

Синонимы эффективно используются для соединения смежных предложений в тексте. Учитывая эти возможности, в учебном корпусе узбекского языка был создан синонимайзер. Синонимайзер – это программа, которая заменяет фразы или слова в тексте синонимами (находящимися в базе данных программы)³¹. В некоторых источниках его называют синонимизатором, синониматором. Этот термин впервые был придуман копирайтером по имени Кевиндарк на форуме SEO для описания программы, которая может упростить переписывание текстов.

Синонимайзер служит для изменения уникальности текста. Это может быть программа-синонимизатор или веб-сайт в Интернете. В мировой компьютерной лингвистике синонимизаторы в основном используются при переписывании (рерайтинг) и копирайтинге для создания уникального контента для веб-сайтов.

Принцип работы синонимайзеров

Синонимайзеры можно разделить на автоматические и полуавтоматические. Полуавтоматические синонимайзеры предлагают пользователю выбрать подходящий из списка синонимов, соответствующий слову, которое он определил. Полностью автоматические синонимы автоматически заменяют слова в тексте значениями используемых символов.

Существует два типа текстовых заменителей, специфичных для автоматических синонимизаторов: неморфологический (линейный) и морфологический (лексический)³².

При неморфологической (линейной) замене учитывается программный бэкэнд заданного слова в тексте, т. е. процесс замены в базе данных слова, содержащего с обеих сторон неисчислимы символы букв, на его аналог по значению, соответствующий той же форме (Абжалова, 2021, с. 12-17). Точнее говоря, этот процесс опирается только на толковый словарь синонимов (Ходжиев, 1974; Отабеков, Бойматова, 2007, с. 49) и вспомогательные словари (Толковый словарь, 2006; Бободжонов, Исламов,

²⁸ <http://rusolimp.kopeisk.ru/leksik/?file=111>; синонимы неполные – это... Что такое синонимы неполные? (academic.ru)

²⁹ <https://synonymonline.ru/synonyms.html>

³⁰ <https://ru.wiktionary.org/wiki/quasi>

³¹ <https://xn----7sbbaqhlkm9ah9aiq.net/news-new/sinonimizator-ru.html>

³² <https://animatika.ru/info/gloss/synonymizer.html#:~:text=>

2007, с. 38; Курбанова, Абжалова и др. 2021, с. 988). Форма слова не учитывается. Например, вместо слова “агар” предложены его синонимы “мабодо”, “магар”, “башарти”, “бордию” (см. рис. 1).

The screenshot shows the 'O'zbek tili milliy korpusi' interface. At the top are buttons for 'Bosh sahifa', 'Lug'atlar', and 'Kirish'. A search bar contains the word 'agar' with an 'Izlash' button. To the right, there are checkboxes for 'Izohli lug'at', 'Omonimlar lug'ati', 'Sinonimlar lug'ati' (checked), 'Antonimlar lug'ati' (checked), and 'Paronimlar lug'ati'. Below the search bar, the results are displayed: 'So'z tarkibi: [agar]', 'Antonim(lar): mavjud emas.', and 'Sinonim(lar): mabodo [f-t], magar [f-t], basharti [f-t+a], bordiyu'.

Рис. 1. В синонимайзере представлены синонимы слова «агар» («если»).

Picture. 1. The synonymizer presents synonyms for the word "agar" ("if").

Однако, если вы добавите аффикс местного падежа и наберете словоформу “агарда” в строке поиска, информация не будет предоставлена (см. рис. 2).

The screenshot shows the same interface as Figure 1, but with the word 'agarda' in the search bar. The results are: 'So'z tarkibi: [agarda]', 'Antonim(lar): mavjud emas.', and 'Sinonim(lar): mavjud emas.'.

Рис. 2. Синонимизатор не может предоставить синонимы слова «агарда» («а если»).

Picture 2. The synonymizer cannot provide synonyms for the word "agarda" ("what if")

Морфологические (лексические) замены также учитывают грамматическую форму слова, и любая словоформа заменяется соответствующей ей формой. Некоторые программы могут изменять не только слова, но и фразы длиной до десяти слов. Некоторые синониматоры могут выполнять оба типа замены. Если программа не может различить морфологию, в результате возможности леммирования или стем различных словоформ в базе данных вручную или с помощью различных утилит, появляется возможность представления синонимов к словоформам или сопоставления их по смыслу. Неморфологический тип синонимайзера был запущен в Национальном корпусе узбекского языка (см. рис. 3), который является первым синонимайзером, созданным для узбекского языка (Абжалова, 2021, с. 38-41). Ведется работа по совершенствованию базы данных, чтобы можно было заменять словоформы по их морфологическому типу. На рисунке 3 словом «катта», введенное в поисковую строку столбца «Словари» Национального корпуса узбекского языка, были указаны синонимы по корпусу и даже слова с противоречивыми значениями (антонимы).



Рис. 3. Неморфологический тип синонимайзера в Национальном корпусе узбекского языка.

Picture 3. Non-morphological type of synonymizer in the Uzbek National Corpus.

Проблема с синонимайзерами. Основные проблемы заключаются в следующем:

- выбор синонимов должен зависеть от контекста;
- инверсия предложения усложняет работу синтаксического анализатора и, как следствие, теряется возможность замены словосочетаний их синонимами;
- если новые слова (термины) не включены в базу данных, неологизмы не будут представлены пользователю.

Чтобы синонимизированный текстовый стиль был свободным, программа должна учитывать закон Зипфа, то есть активные слова должны быть заменены активными синонимами, а редкие слова должны быть заменены редкими синонимами.

Заключение

Обработка синонимизированного текста требует дополнительной грамматической, синтаксической и семантической обработки в других лингвистических компьютерных программах. В процессе обработки тексты сравниваются между собой, после чего выбирается наиболее подходящий вариант.

Литература

Абжалова, М. Вопрос создания синонимайзера или синонимизатора в Национальном Корпусе узбекского языка // Теоретические и практические вопросы создания узбекского национально-просветительского корпуса / Материалы Международной научно-практической конференции. 7 мая 2021 года. Электронное издание / электронная книга. Ташкент : ТошГУУЯЛ, 2021. С. 38-41.

Абжалова, М. А. Возможности поиска словоформ в национальном корпусе узбекского языка на основе лексикографической базы. // Компьютерная лингвистика : проблемы, решения, перспективы / Материалы республиканской научно-технической конференции. Электронное издание. Ташкент : Ташкентский государственный университет узбекского языка и литературы, 23.04.2021. С.12-17.

Бободжонов, Ш., Исламов, И. Словарный классификатор узбекского языка : для школьников. Ташкент : Янги аср авлоди, 2007. 38 с.

Корсакова, Ю. С. Синонимия – универсальное языковое явление // Филологические науки. Вопросы теории и практики, № 1 (5) 2010, часть 2. Тамбов, 2010. С. 103-106.

Курбанова, М., Абжалова, М. [и др.]. Акцентный словарь заимствованных слов узбекского языка : учебно-методический словарь. Ташкент : Нодирабегим, 2021. 988 с.

Отабеков, Ш., Бойматова, Б. Словарь синонимов узбекского языка : для школьников. Ташкент : Янги аср авлоди, 2007. С. 49.

Толковый словарь узбекского языка : более 80 000 слов и словосочетаний (под ред. А. Мадвалиева).

5 томов. Ташкент : Национальная энциклопедия Узбекистана, 2006.

Ходжиев, А. Толковый словарь синонимов узбекского языка. Ташкент : Учитель, 1974. 308 с.

References

Abjalova, M. The issue of creating a synonymizer or synonymizer in the National Corpus of the Uzbek language // Theoretical and practical issues of creating an Uzbek national educational corpus / Materials of the International Scientific and Practical Conference. May 7, 2021. Electronic edition / e-book. Tashkent : TSUULL, 2021 . pp. 38-41. (in Russ.).

Abjalova, M. Possibilities of searching for word forms in the national corpus of the Uzbek language based on the lexicographic base. // Computational linguistics: problems, solutions, prospects / Materials of the republican scientific and technical conference. Electronic edition. Tashkent : Tashkent State University of Uzbek Language and Literature, 04/23/2021. pp.12-17. (in Russ.).

Bobojonov, Sh., Islamov I. Dictionary classifier of the Uzbek language: For schoolchildren. – Tashkent: Yangi asr avlodi, 2007. – 38 p.

Korsakova, Yu. S. Sinonimiya - universal'noe yazy'kovoe yavlenie // Filologicheskie nauki. Voprosy` teorii i praktiki, № 1 (5) 2010, chast` 2. Tambov, 2010. pp.103-106. (in Russ.).

Kurbanova, M., Abjalova M. and others. Accent dictionary of borrowed words of the Uzbek language: [Text]. educational-methodical dictionary / M. Kurbanova, M. Abdzhalova, N. Akhmedov, R. Tulaboeva. Tashkent : Nodirabegim, 2021. 988 p. (in Russ.).

Otabekov Sh., Boymatova B. Dictionary of synonyms of the Uzbek language: For schoolchildren. Tashkent : Yangi asr avlodi, 2007. 49 p. (in Russ.).

Explanatory dictionary of the Uzbek language: more than 80,000 words and phrases (edited by A. Madvaliev). 5 volumes. Tashkent : National Encyclopedia of Uzbekistan, 2006. (in Russ.).

Khodzhiyev A. Explanatory dictionary of synonyms of the Uzbek language. Tashkent : Teacher, 1974. 308 p. (in Russ.).

Для цитирования:

Абжалова М. А. Синонимайзер (синонимизатор) в образовательном корпусе узбекского языка [Электронный ресурс] // Родные языки и культуры в современном изменяющемся мире. 2022. №1. URL.: (дата обращения: дд.мм.гг.). DOI:

For citation:

Abjalova M. A. Synonymizer in the educational corpus of the Uzbek language. Native languages and cultures in the modern changing world. 2022. №1. [online]. Available at (access date ...). DOI:

Дата поступления/Submission date: 06.09.2021.