



**«I International Book Edition
of the countries of the Commonwealth
of Independent States
«BEST YOUNG SCIENTIST – 2020»**

**«I Международное книжное издание
стран Содружества Независимых Государств
«ЛУЧШИЙ МОЛОДОЙ УЧЕНЫЙ – 2020»**



г. Нур-Султан, Казахстан, 13-17 марта 2020г.



поверхности воды, взвешенных в толще слоя, и отложения их на дне; сокращение в воде количества растворенного кислорода вследствие расхода его на окисление поступающих в водоем органических загрязняющих веществ; появление новых бактерий, в том числе и болезнетворных [4].

В целом, состояние и степень экологической безопасности гидросфера определяются экологически безопасными нормами загрязнений объектов. Откуда следует что, качество воды должно отвечать требованиям, предъявляемым организацией экологического нормирования допускаемых антропогенных изменений гидросферы. Однако, зачастую концентрация вредных веществ в сточных водах значительно выше ПДК, установленных СНиП II-32-74 для этих вод при сбросе в городскую канализационную сеть и водоемы.

Долгое время все промышленные отходы и стоки сбрасывались в водные системы или поступали в почву, без какой бы то ни было очистки и контроля. Считалось, что природа своими силами способна переработать опасные отходы процессами очищения. Однако в последние годы участились случаи отрицательного воздействия загрязнения на здоровье людей и окружающую среду в результате техногенных сбросов сточных вод предприятий, превышающих нормы ПДК.

Анализируя вышесказанное, можно сделать вывод, что разработка и внедрение новой технологии, которая обеспечит хорошее качество произведенной мытой шерсти и не даст побочных отрицательных эффектов, является целесообразной и своевременной.

Список литературы

1. Зайцев В.А. Промышленная экология.-М.:РХТУ им. Д.И. Менделеева, 2000
2. Воздействие на организм человека опасных и вредных экологических факторов. Метрологические аспекты (под.ред. Исаева Л.К.). Т.1. - М.: Высшая школа.- 1997. -510 с.
3. Экология и безопасность. Т. 2. Экологическая безопасность Ч.1. (под.ред. Рыбальского Н.Г.). М.: ВНИИПИ.-1993. -300 с.
4. Ласкорин В.М., и др. Проблемы развития безотходных производств. -М.: 2008 г.-56 с.

ТЕХНОЛОГИЯ АНАЛИЗА ОМОНИМОВ ДЛЯ ПРОГРАММ АНАЛИЗА ТЕКСТОВ НА УЗБЕКСКОМ ЯЗЫКЕ

Манзура Абжалова Абдурашетовна

Доктор философских наук (PhD)

*Навоийского государственного горного института
Узбекистан.*

Аннотация: В данной статье рассматривается технология анализа омонимов, встречающихся в узбекских текстах, с помощью программы автоматического редактирования и анализа текстов на узбекском языке. Омонимы, встречающиеся в тексте, изучается отдельно в русском и зарубежном языкознании в зависимости от языкового корпуса. Поскольку узбекский национальный корпус не был создан, эта технология была разработана для лингвистических программ. Этот технологический процесс считается целесообразным применять на языках, не имеющих национального корпуса.

Ключевые слова: лингвистическое программное обеспечение, омоним, корпус, модель, тег, идентификатор.

Abstract: this article discusses the technology for analyzing homonyms found in Uzbek texts using the program for automatic editing and analysis of texts in the Uzbek language. Homonyms found in the text are studied separately in Russian and foreign linguistics, depending



on the language corpus. Since the Uzbek national corpus was not created, this technology was developed for linguistic programs. This technological process is considered appropriate to use in languages that do not have a national corpus.

Keywords: linguistic software, homonym, corpus, model, tag, identifier.

Известно, что написание омонимов принято считать формами, состоящими из одних и тех же фонем. Поэтому в омонимии план выражения является ведущим [1]. Как во многих языках на узбекском языке встречаются абсолютные омонимы (*айри* – **прилагательное**: развилина, раздвоенный; *айри* – существительное: развилка, подставка), частичные омонимы (*гина* – существительное: неприязнь; -*гина* – частица: “*сенигина*” – лишь ты), грамматические омонимы или омоформы (*терим* “тери+м” – существительное: «моя кожа» и *терим* – существительное: «урожай»).

Поскольку не существует универсальной системы (корпуса), охватывающей морфологический уровень языка, в сочетании с аффиксами слов в рамках лексики или словарного запаса, которые выделяются во многих языках, проводятся с морфологическими словарями,ключенными в лингвистический базу. Поэтому понятие омонима (неопределенности) для разных словарей может незначительно отличаться.

Помимо морфологических словарей существует полный набор текстов, которые называются корпусами [2]. В зависимости от формирования корпусного предложения проявляются не только морфологические, но и семантические и синтаксические свойства омонимических форм [3]. В результате создания удобных возможностей для поиска информации корпус становится отличным источником информации о языке, например, можно узнать статистику наиболее распространенных типов омонимических форм.

Сегодня языковой корпус для лингвистических исследований и решения практических задач стал неоспоримым рабочим оружием современной лингвистики, поскольку корпус служит для решения различных лингвистических задач [4,5,6].

Поскольку корпус обрабатывается на основе лексических, морфологических, грамматических, семантических признаков к определенному электронному источнику, он предоставляет очень удобную возможность для использования в лингвистических исследованиях. Ш.Хамроева отметила, что языковой корпус предполагает сбор необходимых, полезных и интересных текстов для изучения, изучения и преподавания определенных текстов в отличие от электронной библиотеки [5].

У разных корпусов явления омонимов различны. Это зависит не только от использования разнообразных моделей морфологии, но также от стиля и объема текста. Тексты, объединенные в книгу русского языка, содержат много общих фраз. Именно поэтому вопрос об устранении омонимии (“снятия омонимии”) в зарубежной и русской лингвистике изучается отдельно [6,7,8].

Из грамматических источников ясно, что почти все методы устранения омонимии (то есть устранения неопределенности) делятся на две группы [8]:

1. Грамматические методы. Они сгруппированы следующим образом:
 - ручные методы без вмешательства технологии;
 - методы автоматического создания правил.
2. Статистические методы.

Каждая из этих групп имеет свои преимущества и недостатки. Могут быть случаи, когда комбинация методов из обеих групп может дать лучший результат. Стоит отметить, что в упомянутых выше исследованиях явления омонимов изучался в связи с корпусной лингвистикой.

Однако Национальный корпус многих языков еще не создан. Соответственно, с такими методами, как метод Бриля, скрытая марковская модель и модификации модели, не будет опоры при анализе омонимов в тексте языков, не имеющих Национального корпуса.



Учитывая тот факт, что узбекский национальный корпус не создан, в связи с N-gammat в анализе форм омонимов в узбекском языке была разработана технология анализа омонимов. С этой целью омонимы в узбекском языке были введены в лингвистическое обеспечение с идентификатором, что дает возможность проанализировать грамматическое образование омонимов в контексте. Для повышения точности анализа была определена категория омонима и придано ему символическое значение. После этого база данных омонимов (List of homonym) была привязана к модели словосочетаний (Models of CompWords): List of homonym=> Models of CompWords. В результате, в зависимости от N – слова, которое идет с M – омонимом, M – омоним получает Q – аффикс, соответствующие идентификационному номеру (ID) M – омонима, которой введен в лингвистическую базу. То есть M[id]+Q[id].

Мы проанализируем этот метод на примере: *енг₁*→N/существительное (часть одежды, покрывающая руку), *енг₂* →V/глагол (*победить, побеждать*), *енг₃* →V/глагол (*еиль*) омонимы в тексте анализируются с помощью грамматического средства, которое в синтаксической связи со словом, идущим за ним (биграмма – последовательность из двух символов), в соответствии с символами категорий слов в модели словосочетаний, созданной в синтаксическом модуле, принимается. Поэтому здесь мы обратимся к моделям, связанным с N и V: N_{ком-ни}+ V (*енгини кесмоқ*), N_{ком-ра}+ V (*енгига қадамоқ*), Pron_{ни(дир)} + V (*ҳаммани енгмоқ*), Pron_{дан(дир)} + V (*мендан енгилмоқ*), Adj_{дан} + V (Vh) (*кучиздан енгилмоқ*) и др.

В целом метод омонимического анализа объясняется тем, что конкретный омоним связан со словом определенным грамматическим показателем, и этот процесс повышает как возможность анализа, так и точность анализа.

Чтобы исключить омонимию, необходимо "классифицировать" каждое слово, то есть его можно сравнить с леммой – фрагментом предложения и набором морфологических признаков, которые для удобства добавляются в один тег. Чтобы изучить все возможные теги, достаточно найти соответствующие ссылки на слова в морфологическом словаре или воспользоваться морфологическим анализатором типа MyStem, который поможет в поиске тегов слов. После этого необходимо будет выбрать только соответствующий тег между несколькими тегами. Оптимальный лингвистический метод, используемый при анализе омонимов, станет важным фактором в процессах редактирования и анализа текстов, машинного перевода, обработки текстов.

Использованные литературы:

1. Раҳматуллаев Ш. Ўзбек тили омонимларининг изоҳли лугати (Толковый словарь узбекских омонимов). – Ташкент: Учитель, 1984. – С. 5.
2. www.dialog-21.ru/media/2138/zakharov.pdf Захаров В.П. Корпуса русского языка.
3. Сайт Открытого Русского Национального корпуса – <http://opencorpora.org>.
4. Недошивина Е.В. Программы для работы с корпусами текстов: обзор основных корпусных менеджеров. Учебно-методическое пособие. – СПб, 2006. – С.26.
5. Хамроева Ш. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Фил.фан. бўйича фалсафа докт. (PhD) дисс. автореф. – Қарши, 2018.
6. Кобрицов Б.П. Модели многозначности русской предметной лексики: глобальные и локальные правила разрешения омонимии. Автореф... канд. филол. наук. Москва: РГГУ, 2004.
7. Зеленков Ю.Г., Сегалович И.В., Титов В.А. Вероятностная модель снятия морфологической омонимии на основе нормализующих подстановок и позиций соседних слов (Электрон ресурс). http://www.dialog-21.ru/media/2444/zelenkov_segalovich.pdf
8. Рысаков С.В. Методы борьбы с омонимией. <http://samag.ru/archive/article/3059>.
9. <http://tech.yandex.ru/mystem> – MyStem морфологик таҳлил дастури сайти.