

Министерство образования и науки Российской Федерации  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«Южно-Уральский государственный университет  
(национальный исследовательский университет)»  
Институт лингвистики и международных коммуникаций  
Кафедра лингвистики и перевода

ДОПУСТИТЬ К ЗАЩИТЕ  
Заведующий кафедрой,  
д.филол.н., доцент  
\_\_\_\_\_ /Т.Н. Хомутова/

**ФУНКЦИОНАЛЬНЫЙ ОБЗОР СОВРЕМЕННОГО  
ЛИНГВИСТИЧЕСКОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ  
ОБРАБОТКИ КИТАЙСКОГО ЯЗЫКА**  
ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА  
ЮУРГУ – 45.03.03.2018.286. ВКР

Руководитель: к.филол.н., доцент  
\_\_\_\_\_ /Б.Г.Фаткулин/  
« \_\_ » \_\_\_\_\_ 2018 г.

Автор работы:  
студента группы ЛМ-437  
\_\_\_\_\_ /А.А.Малойкин/  
« \_\_ » \_\_\_\_\_ 2018 г.

Нормконтролёр,  
преподаватель  
\_\_\_\_\_ /О.И. Бабина/  
« \_\_ » \_\_\_\_\_ 2018 г.

Работа защищена с оценкой:  
\_\_\_\_\_  
« \_\_ » \_\_\_\_\_ 2018 г.

Челябинск  
2018

## ОГЛАВЛЕНИЕ

Введение.....	4
Глава 1 Проблемный подход к работе с китайским языком.....	7
1.1 Особенности морфологии китайского языка, вызывающие трудности у русскоязычных китаеведов .....	7
1.1.2 Проблематика знаменательных слов китайского языка: существительное .....	9
1.1.3 Проблематика знаменательных слов китайского языка: глагол .....	11
1.1.4 Проблематика знаменательных слов китайского языка: прилагательное.....	12
1.1.5 Полузнаменательные и полуслужебные слова китайского языка.....	13
1.1.6 Служебные слова китайского языка как грамматическая особенность.....	14
1.2 Синтаксическая составляющая особенностей китайского языка .....	16
1.3 Фонетика китайского языка и трудности в произношении русскоязычных китаеведов .....	20
1.4 Иероглифика, как одна из трудностей перевода .....	21
1.5 Лексикография: Лексическая омонимия и терминология китайского языка.....	22
1.6 Проблематика знаменательных слов китайского языка: числительное .....	26
1.7 Внеязыковые особенности китайского языка .....	27
Выводы по главе 1 .....	28
Глава 2 Обзор программного обеспечения .....	29

2.1 Программное обеспечение, решающее проблемы на морфологическом уровне китайского языка .....	29
2.1.1 Сегментация текстов на китайском языке .....	29
2.1.2 Функция поиска и замены элементов в тексте .....	32
2.1.3 Поиск счётных слов для существительных .....	33
2.1.4 Числительные в китайском языке .....	34
2.2 Программное обеспечение, решающее проблемы на синтаксическом уровне китайского языка .....	36
2.3 Программное обеспечение, решающее проблемы на фонетическом уровне китайского языка .....	37
2.2.1 Конвертация иероглифов в транскрипцию «Пиньинь» и установка тональности. ....	37
2.4 Программное обеспечение, решающее проблемы иероглифики китайского языка.....	39
2.5 Лексикографические проблемы, встречающиеся при изучении китайского языка.....	41
2.5.1 Электронный словарь, как источник поиска неологизмов и терминологии.....	41
2.6 Внеязыковые особенности обработки китайского языка .....	42
2.6.1 Оптическое распознавание текстов .....	43
2.6.2 Китайские единицы измерения и трудности их преобразования .....	44
Выводы по главе 2 .....	45
Заключение .....	47
Библиографический список .....	48
Приложение .....	53

## ВВЕДЕНИЕ

В последнее время наблюдается большой рост потребности в поиске информации в китайских массивах текстов. В частности, в силу расширения кооперационных связей между нашими странами возникает необходимость мониторинга сайтов госзакупок, научных фондов, производителей товаров и услуг. Для продуктивного и успешного использования информации с китайского языка человеку потребуется использовать программное обеспечение для обработки данных. На различных этапах обработки встречаются соответствующие проблемы программного обеспечения, семантические и грамматические исключения и нюансы каждого из обрабатываемых языков, которые требуют незамедлительного решения. Основной и первой особенностью при обработке китайского языка является иероглифическая система языка.

В настоящее время многие лингвистические системы выполняются на разных языках программирования для разных платформ и, таким образом, заведомо несовместимы. При этом, несмотря на то, что среди существующих языков программирования можно выделить ряд более приспособленных к решению лингвистических задач, а также ряд языков, возможности работы с естественными языками которых, значительно расширены дополнительными библиотеками или наборами классов информации, это не делает их в полной мере ориентированными на естественный язык, что ограничивает эффективность и возможности разработчиков. Научно-технический прогресс необратимо и безостановочно идёт вперёд. Прикладная лингвистика играет большую роль в данном направлении и так же, как и любая другая наука имеет место быть среди первых в научно-техническом прогрессе, так как естественный язык имеет свою собственную ветвь и направление в развитии.

В данной дипломной работе будет представлен обзор программного обеспечения для обработки китайского языка, мы представим примеры, а также укажем основные функции программного обеспечения и скрипты, используемые при изучении и написании практической части дипломной работы.

*Актуальность* данной работы обусловлена тем, что на данном этапе развития прикладной лингвистики программное обеспечение, по большей степени, не систематизировано и не представлено широкому кругу начинающих китаеведов. В статьях трудности китайского языка рассмотрены по отдельности и не несут проблемного похода.

*Практическая значимость* данного исследования: Программное обеспечение, предоставленное в качестве старта развития в сфере прикладной лингвистики, поможет китаеведам и лингвистам успешно начать деятельность в сфере обработки китайского языка.

*Теоретическая значимость*: Благодаря теоретической части исследования, особенности китайского языка были интерпретированы как проблемы, с которыми начинающий лингвист-китаевед может встретиться во время обработки китайского языка.

*Предметом исследования* являются проблемы, возникающие у китаеведов при обработке китайского языка. Также программы, используемые для обработки текстов и массивов на китайском языке.

*Объектом исследования* являются лингвистическое программное обеспечение.

*Цель исследования*: Разработка списка инструментария, который способствует решению трудностей китайского языка.

Для достижения поставленной цели в ходе исследования необходимо было решить следующие *задачи*:

1. Выделение трудностей китайского языка вызванные разноструктурностью
2. Нахождение на каждую трудность программного обеспечения для её преодоления
3. Апробировать каждый инструмент и проанализировать результат

*Материалом исследования* является лингвистическое программное обеспечение.

Для достижения поставленных в исследовании задач использовались следующие *методы*: Теоретический анализ литературы, метод сплошной выборки.

*Структура работы*: работа состоит из введения, двух глав, заключения, библиографического списка и приложения.

Во введении обосновывается актуальность темы, определяются объект, предмет исследования, цель и задачи работы.

В первой главе рассматриваются теоретические основы, определения, классификации.

Во второй главе предоставляются основные результаты обработки китайского языка с помощью используемых программ, применяемых скриптов и команд.

В заключении подводятся итоги исследования. Библиографический список насчитывает 42 источника.

# ГЛАВА 1 ПРОБЛЕМНЫЙ ПОДХОД К РАБОТЕ С КИТАЙСКИМ ЯЗЫКОМ

При изучении различных аспектов китайского языка большинство исследователей отмечают, что грамматика не является наиболее сложной его подсистемой. Имеется в виду, что китайский язык, обладая определенными особенностями – отсутствие флексий, отсутствие склонений, спряжений, родов, падежей, иное грамматическое оформление числа и т.д., представляется простым и легко изучаемым. Однако на самом деле и простота грамматики, и легкость в изучении являются вопросом спорным и относительным.

Китайский язык как язык изолирующий характеризуется отсутствием форм словоизменения. Грамматические отношения между словами в предложении выражаются в этом языке порядком слов, служебными словами и интонацией. Следовательно, в китайском языке отсутствует формальное различие между частями речи – существительным, глаголом и прилагательным, и грамматические категории выражены непосредственно синтаксическими связями.

## **1.1 Особенности морфологии китайского языка, вызывающие трудности у русскоязычных китаеведов**

Для более полной характеристики рассматриваемого вопроса были изучены работы выдающихся китаистов и профессоров – В.А. Курдюмова и В.И. Горелова.

Морфология – (形态学- *xíngtài xué*, 词法- *cífǎ*) – грамматическое учение о слове, раздел грамматики, изучающий систему морфологических особенностей и грамматических категорий, присущих словам китайского языка [Горелов 1989, с. 16].

Слово, являясь важнейшей единицей китайского языка, занимая центральное место в его лексической системе, находится в сложных отношениях с такими понятиями, как слог и морфема. В китайском языке слова, так же, как и в русском, делятся на слоги. Однако, роль слога в этих языках неодина-

кова. В русском языке не каждый слог значим, тогда как в китайском языке, отдельный слог наделен значением. Анализируя факты китайского языка, целесообразно выделить *морфему* как минимальную часть слова, наделенную определенным значением, но лишенную синтаксической самостоятельности [Горелов 1989, с.17]. Некоторые морфемы, обычно лишенные самостоятельного функционирования в речи, в определенных синтаксических условиях все-таки употребляются в качестве самостоятельной единицы языка и тем самым обретают статус односложного слова. Например, морфема – 眼 *yǎn* *глаз* не употребляется самостоятельно, но вместе с тем вполне допустимо сказать: 瞎了一只眼 *xiā le yī zhī yǎn* – *Ослеп на один глаз*. Морфема 题 *tí* *тема* тоже не способна функционировать самостоятельно, но, тем не менее, вполне правомерно такое высказывание, как: 这道题不难 *zhè dào tí bù nán* – *Эта тема нетрудная*.

Слова в китайском языке в большинстве случаев не имеют внешних морфологических признаков, по которым их можно было бы отнести к той или иной части речи. Поэтому при определении принадлежности слов к части речи в основном приходится руководствоваться иными критериями, например способностью слова выступать в роли того или иного члена предложения, соединимостью его со словами других разрядов и с теми или иными формальными показателями. Проблема частей речи является одной из наиболее трудных в грамматике китайского языка. Эта проблема всегда интересовала синологов. Многие исследователи грамматики китайского языка, занимаясь построением системы частей речи, постоянно убеждаются, что границы между китайскими словами существуют, но эти границы относительны и не всегда достаточно определены. Поэтому части речи представляют собой не очень четкую и стройную систему [Коротков 1968, с. 74]. Существует мнение, что самым сложным и в силу этого наиболее спорным вопросом, связанным с проблемой частей речи в китайском языке, является вопрос об использовании одного и того же слова в функции нескольких частей



речи. Можно констатировать, что в современном китайском языке определенные группы слов употребляются в разных синтаксических функциях, постоянно допускают различные транспозиции и при этом не изменяют свою внешнюю словообразовательную форму.

В китайской лингвистической литературе словарный состав языка нередко делят на *实词 shící* полные слова и *虚词 xūcí* пустые слова. Вместе с тем существует также деление на *实义词 shí yìcí* знаменательные слова и *辅助词 fùzhùcí* вспомогательные слова. Однако при делении слов на знаменательные и служебные китайские ученые нередко относят к служебным словам наречия, а также некоторые другие знаменательные части речи. Так, например, Люй Шусян и Чжу Дэси включают в число служебных слов такие части речи, как местоимение, наречие и даже некоторые имена числительные. В результате такого деления только существительные, прилагательные и глаголы как слова, обозначающие предметы, признаки и действия, признаются знаменательными, а все остальные слова автоматически попадают в число служебных. Ван Ляо-и, беря за основу деления слов по частям речи степень конкретности выражаемых этими словами понятий, делит словарный состав китайского языка на четыре разряда: знаменательные слова (имя существительное, имя числительное, имя прилагательное, глагол), полужнаменательные слова (наречие), полуслужебные слова (слова-заместители, в том числе местоимения, а также связки) и служебные слова (соединительные слова и модальные частицы) [Ван 1954, с. 58]. Мы, в след за китайским языковедом Ван Ляо-и придерживаемся данной классификации и даем характеристику каждому разряду.

### **1.1.2 Проблематика знаменательных слов китайского языка: существительное**

В результате изучения различных источников мы пришли к выводу, что существительное *名词 míngcí* занимает важное место в системе частей речи китайского языка. Оно имеет предметное значение. Предметность является

смысловой основой существительного как отдельной лексико-грамматической категории. В китайском языке существительное может обозначать как предмет, так и относительный признак предмета. Обозначая предмет, существительное обычно выполняет функцию подлежащего или дополнения. Оно может также выступать в функции именной части составного или сложного сказуемого. Указывая относительный признак предмета, существительное выполняет функцию определения. Функцию же обстоятельства существительное способно выполнять лишь в определенных условиях, да и то часто в таких случаях оно функционально сближается с дополнением, обозначающим косвенный объект [Коротков 1968, с. 128-129].

Существительные в китайском языке не различаются по родам, не изменяются по числам, не склоняются по падежам. Единственным морфологическим показателем числа в китайском языке служит суффикс *men*, который, являясь обязательным для местоимений множественного числа, с существительными употребляется только в особых случаях, например с существительными, обозначающими лиц, но главным образом при обращении. В каждом конкретном случае число существительных уточняется контекстом, т.е. с помощью числительных или других показателей числа. Род существительных при необходимости тоже может быть выражен лексически, например: 学生 *xuésheng* – студент(ка), 男学生 *nánxuésheng* – студент (*nán* мужчина), 女学生 *nǚxuésheng* – студентка (*nǚ* женщина). Множественное число – 学生们 *xuéshengmen* – студенты.

В китайском языке, как и в европейских языках, существительные разделяются на исчисляемые и неисчисляемые существительные. Однако в отличие от европейских языков, где при подсчете исчисляемых предметов не требуются специальные слова типа единицы измерения, в китайском языке существительные исчисляются с помощью специальных служебных слов, называемых *счетными* словами. Вся масса исчисляемых существительных подразделяется на классы – отчасти по внешнему виду предметов, а в основ-

ном просто по языковой традиции, причем все существительные данного класса употребляются с одним и тем же счетным словом. Например, существительное 人 *rén* – человек требует после себя универсальное счетное слово 个 *gè*, которое указывает на нейтральный класс предметов, имеет обобщенное значение – 三个人 *sān gè rén* – три человека или 他们是三个人 *tāmen shì sān gè rén* – их было трое. Существительное 车 *chē* – машина требует после себя счетное слово для повозок и экипажей – 辆 *liàng*. Если рассмотреть класс предметов, которые по форме напоминают продолговатую форму, такие как тонкие, узкие предметы, реки, улицы продолговатые предметы стандартной формы и размера, то мы будем употреблять счетное слово 条 *tiáo*, например 一条大河 *yītiáo dàhé* – одна большая река, 一条大街 *yītiáo dàjiē* – один проспект.

### 1.1.3 Проблематика знаменательных слов китайского языка: глагол

Глагол 动词 *dòngcí* в китайском языке самая сложная и интересная часть речи. Она занимает важное место в системе частей речи китайского языка. Глагол не изменяется по числам и родам, ему не свойственно также спряжение по лицам. Видо-временные значения выражаются с помощью глагольных суффиксов. Суффикс 了 указывает, что действие имело место в определенное время в прошлом и что оно завершено. Суффикс 过 указывает, что единичное или повторявшееся несколько раз действие имело место в неопределенное время в прошлом и что оно завершено до данного момента. По мнению С. Е. Яхонтова, эта глагольная форма обозначает действие, имевшее место в неопределенное время в прошлом и никак не связанное непосредственно с настоящим моментом. Характерно, что временные значения в китайском языке могут передаваться также с помощью отрицаний, различных для разных времен. Показателем временного значения служит также отсутствие суффиксального оформления. Неоформленный глагол передает, как правило, действие, относящееся к настоящему или будущему времени. В китайском языке

отсутствует система залогов и наклонений. Значение, соответствующее страдательному залогу, выражается аналитически, т.е. с помощью специального служебного слова. Единственным наклонением, существующим в китайском языке, является потенциальное наклонение, указывающее на возможность или невозможность совершения действия. Образуется эта форма с помощью инфиксов – аффиксов, вставляемых внутрь слова. В китайском языке имеется богатая система модальных глаголов, выражающих возможность, необходимость или желание совершить действие. Особенностью этой категории глаголов является их полная морфологическая неизменяемость, они не образуют никаких форм – ни синтетических, ни аналитических. Из полнозначных глаголов наибольший интерес представляют результативные глаголы и глаголы направления движения. Многосложные глаголы направления движения характеризуются своеобразной структурой, которая позволяет выражать единым словом движение, ориентированное не только в пространстве (вверх, вниз, внутрь и т.д.), но и к говорящему или от говорящего, например: *跑进去* *pǎo jìnqù* вбежать, *跑出来* *pǎo chū lái* выбежать.

#### **1.1.4 Проблематика знаменательных слов китайского языка:**

##### **прилагательное**

Прилагательное 形容词 *xíngróngcí* в китайском языке, как и существительное, не изменяется по родам, числам и падежам. Прилагательное не имеет форм степеней сравнения, эти значения выражаются в языке лексически. В системе частей речи прилагательное по своим грамматическим свойствам ближе к глаголу, чем к существительному. В то время как в русском языке прилагательные вместе с существительными образуют одну более общую категорию имени, в китайском языке прилагательные объединяются с глаголами в общую категорию предикатива. Хотя, большинство исследователей грамматики китайского языка, в отличие от А.А.Драгунова, не видят достаточных оснований для того, чтобы объединить глагол и прилагательное в одну более общую грамматическую категорию предикатива [Горелов 1989,

с.37]. Как и глаголы, китайские прилагательные самостоятельно, без помощи связки, выступают в роли сказуемого, свободно оформляются видо-временными суффиксами, непосредственно соединяются с отрицанием и т. п. В качестве именного члена предложения (подлежащего, дополнения и именной части сказуемого) прилагательное может выступать, только будучи оформленным морфемой *de*, которая придает ему именной характер.

### 1.1.5 Полузнаменательные и полуслужебные слова китайского языка

Полузнаменательными словами в китайском языке являются наречия, а полуслужебными – местоимения и связки.

Наречие *副词* – часть речи, обозначающая признак действия или признак качества, а также степень признака. Кроме того, наречие может указывать место и время совершения действия или проявления качества. В китайском языке наречие способно обозначать также относительный признак предмета. Наречие, таким образом, выражает многообразные значения, имеет широкую и сложную семантическую основу. В соответствии с общим грамматическим значением и формально-грамматическими свойствами, и особенностями наречия в китайском языке подразделяются на три класса: качественные наречия, обстоятельственные наречия и наречия степени. [Горелов 1989, с.63].

Качественные наречия отражают внутреннюю сторону развития действия, содержат качественную характеристику действия. *忽然* *вдруг*, *连忙* *постепенно* и другие.

Обстоятельственные наречия отражают внешнюю сторону развития действия или проявления признака. Они указывают обстоятельства, при которых совершается действие или проявляется признак. *已经* *уже*, *也* *также*, *马上* *тотчас* и др. Наречия данного класса сочетаются с глаголами, а также с прилагательными.

Наречия степени указывают степень проявления признака, обозначают степень качества или интенсивность действия. *更* *еще более*, *级* *крайне* и др.

Наречия этого класса в большинстве случаев употребляются с прилагательными.

Что касается местоимений, то они, как и существительные и прилагательные не изменяются по падежам, не имеют категории рода и числа. По значению местоимения китайского языка подразделяются на личные, притяжательные, указательные, определительные, вопросительные, неопределенные и относительные. В китайском языке нет отрицательных местоимений. Значения, передаваемые местоимениями этого класса в русском языке, получают свое выражение в китайском языке при помощи неопределенных местоимений, усилительных и отрицательных частиц.

Еще один разряд полуслужебных слов китайского языка – связки. Это специальное слово для связи подлежащего и сказуемого. Часто связкой является глагол. В китайском языке глагол-связка 是 *shì* «быть» часто используется в предложениях, которые строятся по следующей схеме: *подлежащее + глагол-связка 是 + часть (существительное или прилагательное)*. В китайском языке глагол-связка употребляется во всех временах.

### **1.1.6 Служебные слова китайского языка как грамматическая особенность**

Служебными словами в китайском языке принято считать модальные частицы и соединительные слова – союзы и предлоги. Служебные слова употребляются для связи слов, предложений или частей предложений.

Наиболее употребительными для китайского языка являются следующие семь модальных частиц: 啊 *a*, 吧 *ba*, 的 *de*, 了 *le*, 吗 *ma*, 呢 *ne*, 呀 *ya*, хотя, частицу 呀 принято считать фонетической разновидностью частицы 啊. Значение частицы 啊 – обозначение различных чувств и эмоций в восклицательных предложениях; для частицы 吧 – выражение неуверенности в общих вопросах и условных предложениях, а также обозначение согласия и побуждения; для частицы 的 – усиление утверждения, часто реализуемое в двух типах вы-

делительной конструкции 是.....的; для частицы 了 – указание на изменение ситуации или состояния; для частицы 吗 – обозначение общего вопроса; для частицы 呢 – обозначение специального и альтернативного вопросов [Симатова 2014, с.33].

Что касается союзов, то они образуют самый многочисленный класс служебных слов. Их значение как средства синтаксической связи весьма велико. Особенно значительна роль союзов в произведениях письменнокнижных стилей. В китайском языке многие союзы происходят от глаголов. Кроме союзов глагольного происхождения, имеется также некоторое число союзов наречного и местоименного происхождения. Союз имеет значение не сам по себе, а как выразитель того или иного сочетания, как словесное обнаружение такого сочетания. [Шахматов 1927, с.98]. По значению союзы китайского языка подразделяются на девять разрядов, каждый из которых обладает своими семантико-функциональными особенностями. Соединительные союзы: 和 *hé* и, 而且 *érqiě* к тому же. Противительные союзы: 但是 *dànshì* но, 然而 *rán'ér* однако. Разделительные союзы: 或者 *huòzhě* или, 还是 *háishi* или. Сопоставительные союзы: 如果...则 *rúguǒ... zé* если, то. Целевые союзы: 为了 *wèile* для того, чтобы. Причинные союзы: 因为 *yīnwèi* по причине. Условные союзы: 要是 *yàoshi* если. Уступительные союзы: 虽然 *suīrán* хотя. Сравнительные союзы: 像 *xiàng* как. Иногда в китайском языке функцию союзов выполняют сочетания предлогов с послелогоми. Эти сочетания употребляются в качестве союзов в сложноподчиненных предложениях, выражающих временные отношения, а также в предложениях с придаточной частью, выражающих эти же отношения [Горелов 1963, с.77]. В роли союзов используются следующие сочетания предлогов с послелогоми: 在 ... 以前 *zài ... yǐqián* до того как, 在... 以后 *zài ... yǐhòu* после того как и др.

Один из важнейших классов в системе служебных слов многих языков образуют предлоги, и китайский язык не является исключением. Китайские

лингвисты отмечают, в отдельных, единичных случаях предлоги, подобно глаголам, способны образовать в предложении вопрос путем повтора в положительной и отрицательной форме. Предлоги китайского языка в своем большинстве существуют рядом с глаголами, от которых они произошли: *在* находится и *不在* в, на; *到* прийти, приехать и *不到* до, к; *向* быть обращенным к и *不向* к, напротив. В китайском языке по вопросу о соотношении предлогов и глаголов, от которых произошли предлоги, существуют различные точки зрения. Так Ли Цзиньси и Лю Шыжу, а также Люй Шусян считают предлоги служебными словами и полагают, что они не могут одновременно выступать как глаголы и, следовательно, выполнять функцию сказуемого. Таким образом, они признают невозможность совмещения в одном слове знаменательной и служебной функции. Другая точка зрения допускает существование в китайском языке так называемых бифункциональных слов, совмещающих знаменательную и служебную функции, в данном конкретном случае – функции глагола и предлога [Янквер 1964, с. 5].

Таким образом, теоретический анализ литературы показывает, что проблема морфологии китайского языка рассматривалась достаточно широко как русскими лингвистами, так и китайскими.

## **1.2 Синтаксическая составляющая особенностей китайского языка**

Проблемой изучения синтаксиса китайского языка занимались такие российские ученые как: Шутова Е.И., Курдюмов В.А., Горелов В.И., Солнцев В.М., Солнцева Н.В., Драгунов А.А. и зарубежные исследователи: Ван Ли, Чжао Юньпин, Цзинь Шаочжи, Ху Мин-ян и др.

В последние годы появились или были серьёзно переработаны такие значительные труды, как «Курс китайского языка. Теоретическая грамматика» В.А. Курдюмова, «Морфология результативных конструкций» К.В. Антонян, «Введение в китайский язык» Н.А. Спешнева, работы Н.А. Деминой, И.В. Кочергина, О.А. Масловец, М.Г. Фроловой, О.М. Готлиба и многие другие [Кошкин 2013, с.106].



Синтаксис китайского языка характеризуется строгой системой правил, касающихся порядка слов, соединения слов в предложении и структурных особенностей предложений различного типа. В китайском языке порядок слов в предложении твердо определен [Курдюмов 2006, с.76]. Грамматическая роль слова и его связи с другими словами в предложении устанавливаются главным образом за счет порядка слов, нарушение которого приводит к изменению грамматической роли и связи слов. Наиболее распространенная форма простого предложения соответствует схеме (рис 1.1).



Рисунок 1.1 – Графическая модель предложения

Это основополагающая модель китайского предложения, являющаяся исходным материалом для образования других конструкций. Возможные перестановки слов в китайском предложении тоже имеют различного рода формально-грамматическую обусловленность. Характерно, что синтаксические отношения внутри словосочетаний, а также аналогично построенных сложных слов складываются только в соответствии с основной, исходной моделью: подлежащее всегда предшествует сказуемому, дополнение всегда следует за сказуемым. Схема «подлежащее – сказуемое – дополнение» едина для всех коммуникативных типов предложения, в том числе и для вопросительного. Общий вопрос выражается в китайском языке не изменением порядка слов, как в английском, немецком и других языках, а с помощью вопросительных частиц, замыкающих предложение, или путем особого повтора сказуемого: глагол или прилагательное, выступающее в роли сказуемого, дублируются в положительной и отрицательной

форме, что в буквальном переводе дает построения типа: «пойдешь – не пойдешь?», «читал – не читал?» и т. п. Последний способ образования вопроса является очень распространенным. Отсутствие в китайском языке падежных окончаний усиливает и грамматическую роль служебных слов, указывающих на связь и отношения между словами. Служебные слова являются важнейшими ориентирами при чтении и слушании китайского языкового материала. Своеобразное явление представляют собой предлоги китайского языка. Все они произошли от глаголов и в той или иной степени сохраняют связь с лексическим значением соответствующих глаголов, что дает основание называть их глаголами-предлогами. Многие из глаголов, от которых произошли предлоги, продолжают широко употребляться в современном китайском языке. Таким образом, в языке существует немало парных элементов, одинаково звучащих и записывающихся одним и тем же иероглифом, но выполняющих разные грамматические функции: глагол – полнозначную сказуемостную функцию, предлог – служебную, выражая при этом объектные, пространственные, временные, целевые и причинные отношения.

Китайский язык имеет развитую систему сложных предложений, образуемых союзным и бессоюзным сочинением и подчинением. Специфичной для китайского языка разновидностью сложных предложений являются последовательно связанные предложения, или, как их еще называют, «предложения сцепления». В предложениях такого типа имеется один общий член, выполняющий двоякую синтаксическую функцию: например, он является дополнением к сказуемому начальной части предложения и одновременно служит подлежащим в отношении сказуемого последующей части предложения. Входя в состав обеих частей предложения, такое слово является как бы связующим звеном между ними. Особую разновидность составляют сложные предложения с включенной частью, в которых тот или иной член, будучи распространенным до целого предложения, входит в состав сложного предложения непосредственно, без какого-либо грамматического оформления в

виде союзов или других служебных слов. Характерно, что порядок слов внутри такого распространенного члена предложения строго определен, он всегда соответствует основной, исходной структуре.

Опишем основные синтаксические единицы китайского языка – предложение и словосочетание. Для словосочетаний и предложений китайского языка характерны структурная простота, стройность и четкость внутренней организации. Китайский лингвист Линь Юйвэнь следующим образом характеризует свойства и особенности синтаксических построений в китайском языке. Абсолютное большинство слов китайского языка – однослоги и двуслоги. Это создает возможность использования в полной мере синтаксических структур, обладающих стройной и симметричной организацией, и синтаксических структур, для которых характерно переплетение и скрещивание компонентов. Синтаксические структуры первого типа не только могут сочетаться между собой, но вместе с тем могут также объединяться с синтаксическими построениями второго типа, тем самым обогащая и делая многообразным китайский язык [Люй 1965].

В китайском языке формально-грамматическими средствами обозначения синтаксических связей и вместе с тем выражения смысловых отношений между компонентами словосочетаний являются служебные слова и словопорядок. В китайском языке порядок слов в структуре словосочетаний выполняет важную грамматическую функцию. Наглядно это можно показать, сопоставив следующие словосочетания: *保护环境 охранять окружающую среду и 环境保护 охрана окружающей среды.*

При научном описании словосочетаний немаловажную роль играют вопросы классификации. Классификация позволит привести в систему все многообразие словосочетаний и наглядно показать характерные черты и особенности синтаксических единиц данного класса. Распределяя словосочетания в соответствии с присущими им структурными и семантическими признаками, целесообразно дать двойную классификацию: структурную – по числу ком-

понентов (знаменательных слов); семантическую – по смысловой слитности компонентов, а также по центральному компоненту (в зависимости от его категориальной принадлежности) [Горелов 1989 с. 103].

Описывая предложение китайского языка нужно сказать, что предложения неоднородны. Для них характерна разная коммуникативная установка. Наличие или отсутствие эмоциональной окраски также представляет собой один из дифференциальных признаков китайских предложений. В зависимости от функции, цели высказывания, иногда называемой коммуникативной установкой, предложения китайского языка подразделяются на *повествовательные, вопросительные и побудительные*. Обычным для повествовательных предложений является употребление в их составе экспрессивных (фразовых) частиц: 吧, 了, 呢. Например: *晚场电影八点才开呢* *Ведь вечерний киносеанс начнется только в 8 часов.*

Вопросительные предложения подразделяют на предложения, содержащие специальный, общий, а также альтернативный вопрос. С функциональной точки зрения побудительные предложения существенно отличаются от предложений повествовательных. Д.А. Штелинг обращает внимание на глубокое, принципиальное различие, существующее между высказыванием, содержащим информацию о действительности, и высказыванием неинформативным, в котором говорящий проявляет свою волю [Штелинг 1973, с. 66].

### **1.3 Фонетика китайского языка и трудности в произношении русскоязычных китаеведов**

С фонетической точки зрения звуковой поток, или речь, распадается по слуху на минимальные единицы, которые мы называем слогами. При аналитическом рассмотрении слога в нем можно выделить отдельные звуки. Структура слога в китайском языке заметно отличается от структуры слога в русском языке. В русском языке слог может состоять из разного количества звуков. В китайском же языке количество звуков в слоге не превышает четырех. [Спешнев 1980, с. 8]. Отдельные звуки, входящие в состав китайского

слога, размещаются в нем в строгом порядке. Каждый звук занимает только отведенное ему место. В пределах одного китайского слога, например, невозможно сочетание согласных, зато возможно сочетание гласных, чего не бывает в слогах русского языка.

В китайском языке отсутствует алфавит. В алфавите буквы обозначают звуки. В Китае же письменные знаки обозначают не звуки, а определенные морфемы языка. Каждая морфема имеет свое значение. Такой вид письма называется логографическим [Ильницкая 2015, с. 89]. Долгое время было непонятно, как записывать произношение китайских иероглифов и в 1958 г. была принята система пиньинь (拼音). В системе пиньинь все звуки можно разбить на инициали – первая часть слога и финали – завершающая часть. Всего существует 21 инициаль и 36 финалей. В системе пиньинь используют буквы латинского алфавита. Произвольное сочетание инициалей и финалей недопустимо, есть определенный набор между ними. В путунхуа 普通话 (досл. «общепринятый язык») – стандартном китайском языке – 4 тона. Тоны выполняют смыслообразующую роль. Это значит, что слово, сказанное другим тоном, может приобрести новый смысл, пример: *好酷 hǎo kù* – Очень круто, *好苦 Hǎo kǔ* – Очень горько, *嚎哭 Háo kū* – Громко плакать. Поскольку в китайском языке довольно часто встречаются омонимы, то именно тоны позволяют отличать их друг от друга.

#### **1.4 Иероглифика, как одна из трудностей перевода**

Китайская иероглифическая письменность – исключительное явление среди современных письменностей. Это единственная иероглифическая письменность мира, которая была изобретена за полтора тысячелетия до н. э. и продолжает существовать в наши дни. Китайская иероглифическая письменность на протяжении всей своей истории приспособлялась к меняющимся условиям развития цивилизации и сейчас продолжает оставаться сложным, но приемлемым для Китая средством письма.

В Китае существует два вида китайской письменности: традиционный (繁体字 *fán tǐ zì*, *фаньтицзы*) и упрощенный (简体字 *jiǎn tǐ zì*, *цзяньтицзы*). Упрощенная китайская система написания отличается от традиционной лишь двумя аспектами: сокращенное число штрихов в иероглифе (например, 觀 → 观); сокращенное количество используемых знаков (два разных слова теперь пишутся одним и тем же иероглифом).

Буквально до недавнего времени китайцы писали иероглифы сверху вниз, а столбцы шли справа налево. Сейчас в Китае преимущественно пишут горизонтально, слева направо, как в европейских языках; вертикальное письмо продолжает использоваться на Тайване наряду с горизонтальным. Однако в материковом Китае вертикальное письмо и дореформенная иероглифика (традиционные, полные формы иероглифов) по-прежнему используются как отсылка к традиционной китайской культуре.

### **1.5 Лексикография: Лексическая омонимия и терминология китайского языка**

Лексическими омонимами в китайском языке называются слова, совпадающие по звучанию и иногда по написанию, но различающиеся по значению: 同音词 *tóngyīncí* 同音异义词 *tóngyīnyìyìcí*. Китайский язык один из самых омонимичных языков. По различным подсчетам количество омонимов в лексике составляет от 10% до 16%. В китайском языке омонимия распространена как среди односложных слов, так и среди двусложных, причем односложных омонимов намного больше. С точки зрения отношения к устной или письменной речи можно выделить три типа омонимов. Для китайского языка это очень важно в связи с его иероглифической письменностью.

**Первый тип.** В нем слова совпадают по звучанию (включая инициаль, финаль и тон) и написанию, это собственно омонимы. В этом случае один и тот же иероглиф записывает одинаково звучащие слова. Их называют 同音同形词 *tóngyīn tóngxíngcí* например 词 *cí* имеет значение «слово», «речь» (язы-

ковая единица) и значение 词 cí «литературная форма», название древнего поэтического жанра «цы».

**Второй тип.** В нем слова совпадают по форме в устной речи, но пишутся по-разному, т. е. совпадают по звучанию, но не совпадают по написанию. Это фонетические омонимы, или омофоны. Их называют 同音异形词 *tóngyīn yùxíng cí* или 等音词 *děngyīn cí*. Примеры: 班 *bān* – группа и 斑 *bān* – пятно, пятнистый, 八 *bā* – восемь и 扒 *bā* – выкапывать. В китайском языке наиболее распространен именно этот тип омонимии.

**Третий тип.** В нем слова имеют одинаковое написание, но разное звучание, их называют 同形异音词 *tóngxíng yìyīn cí* – омографы. Примеры: 还 *hái* произносится как 还 *hái* – все еще и 还 *huán* – возвращать, 重 *zhòng* – тяжелый и 重 *chóng* – вновь.

**Источники возникновения омонимов.** Китайская фонетическая система насчитывает всего 420 слогов, в то время как слов в нём несколько десятков тысяч. Процесс образования омонимов в результате фонетических изменений наблюдался на протяжении всей истории китайского языка. Многочисленные процессы, в ходе которых возникали омонимы, можно поделить на 3 группы: Изменение системы финалей. Изменение состава инициалей. Изменение или упрощение тонов. В современном китайском языке существует огромное количество равнозвучащих слов, с помощью которых он выражает то же количество смыслов, что и любой другой язык, имеющий фонетическое разнообразие. В результате произошедших фонетических процессов за одной звуковой оболочкой закрепилось несколько значений. 代 *dài* (эпоха) – 贷 *dài* (занимать) – изначально имели разные инициалы. 天 *tiān* (небо) – 添 *tiān* (увеличить) – ранее имели разные конечные согласные. 后 *hòu* (последний) – 候 *hòu* (ждать) – первоначально были разного тона.

Омонимы возникают также в результате распада полисемии, т.е. деления одного слова на два. В многозначном слове все его значения так или иначе

зависят друг от друга, связаны друг с другом. Может получиться так, что в каком-то месте цепь смысловых зависимостей между значениями рвётся, и тогда появляются слова омонимы. 月 yuè (луна) – 月 yuè (месяц) – возникли в результате переноса значения: от полного цикла фаз луны и времени, которое этот цикл занимает.

Следующий источник омонимов – иностранные заимствования. Так как количество фонетических заимствований в этом языке ограничено, а те, которые всё же попадают, обычно многосложны и не состоят из одинаково звучащих слов, то их роль в китайском языке не так велика. Однако существует немало омонимов в односложной форме, которые передаются по-китайски одним из представленных в китайском языке слов. 米 mǐ (рис) – 米 mǐ (метр).

Диалектная форма также является источником омонимов. Она представляет собой заимствования из диалектов, попавшие в китайский национальный язык. Но наряду с иностранными заимствованиями, диалекты являются источником образования незначительной числа омонимов.

Источником омонимии могут быть различные словообразовательные процессы, например, процесс оформления суффиксом 儿 ér односложных морфем. Присоединяясь к морфеме, происходит эризация слога, суффикс и морфема сливаются. В ряде случаев присоединение суффикса приводит к образованию омонимов, хотя до этого односложные единицы не являлись омонимами. 芽儿 yár – почка, 牙儿 yár – зуб.

Следующий источник омонимов – словообразовательный процесс сокращения слов. 外部 wàibù (МИД) – 外部 wàibù (внешняя сторона).

Данный источник является малопродуктивным для образования омонимов в китайском языке.

Однако бывают случаи, когда омонимия может привести к ошибкам. В большинстве случаев всевозможные недоразумения снимаются благодаря контексту, но бывают случаи, когда омонимия является препятствием для об-



щения. В таком случае в китайском языке используются различные способы для её устранения. В процессе развития современного китайского языка многие односложные слова превратились в двусложные. В процессе превращения из однослогов в двуслоги, многие слова перестают быть омонимами 优良 *yōuliáng* (прекрасный) – 忧愁 *yōuchóu* (тосковать). В китайском языке множество морфем, сходных или близких по значению, которые можно использовать для избегания омонимии 遇见 *yùjiàn* (встретить) – 预见 *yùjiàn* (предвидеть) – в данном случае 遇见 можно заменить на 遇到 *yùdào*. Огромное количество синонимов в китайском языке, также является помощником в устранении синонимии 出版 *chūbǎn* (издавать, печатать) – 初版 *chūbǎn* (первое издание книги) – в данном случае 初版 заменяется на 第一版 *dìyībǎn*.

**Терминология китайского языка.** Терминология современного китайского языка концентрируется преимущественно на специальной лексике и лишь затем – на денотации. Проблемы синонимии и омонимии, обусловленные спецификой китайского языка, являются достаточно непростыми, требующими своего отдельного исследования. Особенностью терминов китайского языка и специальной лексики является специфичность связи ее единиц с понятиями и именуемыми объектами. В частности к проблеме специальной лексики в общей теории терминологии обращаются А. В. Суперанская, Н. В. Подольская, Н. В. Васильева [Суперанская, Подольская, Васильева 2008, с.236].

Известный ученый, изучающий терминологию Б. Н. Головин в своих работах отмечает: «Термин – это отдельное слово или образованное на базе имени существительного, подчинительное словосочетание, обозначающее профессиональное понятие и предназначенное для удовлетворения специфических нужд общения в сфере определенной профессии» [Головин 1966, с. 164].

Все термины китайского языка можно разделить на 3 типа: 1) Простые – состоящие из одного компонента 闸 шлюз; 拖 тяга. 2) Сложные – состоящие

из 2–3 компонентов 采油 нефтедобыча = 采 «сбор, добыча» + 油 «масло»; 石油库 нефтехранилище = 石 «камень» + 油 «масло» + 库 «хранилище, амбар».

3) термины-словосочетания 石油井 нефтяная скважина = 石 «камень» + 油 «нефть» + 井 «колодец». Перевод сложных терминов-словосочетаний всегда осуществляется с существительного, если словосочетание состоит из нескольких существительных, то определяется существительное, которое служит основным компонентом, а затем переводится остальная смысловая группа компонентов. Например, 地下储油库 подземное хранилище нефти = 地 «земля, земная поверхность» + 下 «снизу, под» + 油 «масло» + 库 «амбар, хранилище». Из последнего компонента словосочетания 库 «амбар, хранилище», сразу становится ясным то, что термин означает место хранения, перед компонентом 库 находится компонент 油 «масло, нефть», увидев этот иероглиф, нам становится известным, что хранилище предназначено для хранения нефти.

### **1.6 Проблематика знаменательных слов китайского языка: числительное**

Числительное 数词 shùcí – часть речи, обозначающая отвлеченные числа, количество предметов или порядок предметов при счете. Среди китайских ученых не существует единого мнения по вопросу о том, следует ли выделять числительное в самостоятельную часть речи. Так, например, Люй Шусян и Чжу Дэси считают, что числительное примыкает к более общей категории прилагательного. Они полагают возможным включить некоторые числительные далее в разряд служебных слов. Более правильной точки зрения придерживается Ван Ли. Он считает, что слова, обозначающие числа, по грамматическому употреблению отличаются от прилагательных. Поэтому применительно к китайскому языку предпочтительно признать числительное самостоятельной частью речи [Горелов 1989, с.42].

Числительные в китайском языке подразделяются на два класса: количественные и порядковые числительные. Количественные числительные обозначают отвлеченные числа или количество предметов: 五 wǔ пять, 十五 shíwǔ пятнадцать, 二十年 èrshí nián двадцать лет. Порядковые числительные обозначают порядок предметов при счете. Показателем порядковых числительных является префикс 第 dì – 第二 второй, 第五 – пятый.

### 1.7 Внеязыковые особенности китайского языка

**Единицы измерения в Китае.** На протяжении многих веков единицы измерения различались территориально, что затрудняло развитие торговых отношений между различными регионами Древнего Китая. Территориальная удаленность и обособленность Китая от центров торговли с европейскими и азиатскими странами способствовали формированию китайской системы мер.

В результате контактов Кореи и Японии с Китаем, большинство единиц измерения этих стран стали аналогами китайских мер. Единица длины фэнь первоначально равнялась длине одного зерна проса. Единица длины бу соответствовала двойному шагу. Мера длины чжан равнялась расстоянию между концами пальцев разведенных в стороны рук (сажень).

В Древнем Китае применялась система исчисления нулем и десятичной системой, сформировавшими систему мер и весов.

Современные классические меры длины: 毫米 háomǐ миллиметр, 公分 gōngfēn или 厘米 límǐ сантиметр, 米 mǐ или 公尺 gōngchǐ метр, 公里 gōnglǐ километр. Другие современные меры длины: 英寸 (吋) yīngcùn, cùn дюйм, 英尺 yīngchǐ английский фут. Традиционные китайские меры длины: 寸 cùn цунь, 尺 chǐ чи = 32 см, 里 lǐ ли = 500 м = 0,5 км. Современные классические меры веса: 克 kè или 公分 gōngfēn грамм, 公斤 gōngjīn килограмм, 吨 dūn тонна. Другие современные меры веса: 磅 bàng английский фунт. Традиционные китайские меры веса: 斤 jīn Цзинь = 0,5 кг, 分 fēn Фэнь = 0,5 гр.

## **Выводы по главе 1**

В данной главе нами были рассмотрены проблемы морфологии и синтаксиса китайского языка, фонетики и иероглифики, лексикографии, а также внеязыковые особенности. Грамматические особенности современного китайского языка заметно отличаются от славянских языков. Это обусловлено особенностями культуры, менталитета и тенденцией к экономии грамматических средств, как на уровне синтаксиса, так и морфологии. Кроме этого, можно выделить еще две отличительные черты китайского языка – это фонетика и иероглифика. Китайский язык – это тональный язык, тоны выполняют смысловоразличительную функцию. Что касается иероглифики, то в китайском языке существует как традиционное написание иероглифов, так и упрощенное. Традиционное написание используют в названиях различных объектов, что усложняет чтение и перевод этих иероглифов. Еще одной особенностью китайского языка является омонимия. Поскольку китайский язык является одним из самых омонимичных языков, проблема разграничения омонимов имеет немаловажное значение. На выявленные нами трудности во второй главе будут представлены решения в виде программного инструментария.

## **ГЛАВА 2 ОБЗОР ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ**

Перед началом исследования практической части работы, следует ознакомиться с основными терминами.

Программное обеспечение – совокупность программ системы обработки информации и программных документов, необходимых для эксплуатации этих программ [Селиванов 1989, с. 168].

На сегодняшний день программное обеспечение является одним из основных продуктов новых технологий позволяя использовать весь потенциал развития компьютерных наук.

Обращаясь к практической части данной работы, будут представлены пути решения проблем, представленных в первой главе. Описав основные особенности китайского языка и его проблемы, нельзя оставлять без внимания программное обеспечение, благодаря которому начинающий специалист в сфере прикладной лингвистики изучающий китайский сможет успешно и без проблем перевести текст точно и быстро.

### **2.1 Программное обеспечение, решающее проблемы на морфологическом уровне китайского языка**

Продолжая тему обзора программного обеспечения следует начать с представления инструментов. В данном параграфе будут представлены несколько примеров использования прикладных функций для обработки китайского языка.

#### **2.1.1 Сегментация текстов на китайском языке**

Для дальнейшего исследования следует дать определение термину «Сегментация».

Сегментацией в лингвистике называют линейное членение речевого потока на составляющие отрезки. В науке разделяют 2 разновидности сегментации:

- Звуковой уровень
- Уровень знаковых единиц

Китайский язык, несмотря на кажущуюся сложность, имеет чрезвычайно простую грамматику. Главная проблема при анализе китайских текстов обусловлена двумя особенностями. В китайском языке отсутствуют пробелы между словами, и практически любое сочетание иероглифов может быть интерпретировано тем или иным образом. Это порождает проблему сегментации предложений на слова [Чуцяо 2016, с.138].

На рисунке 2.1 Изображена сегментация текста. Данный инструмент позволит пользователю выполнить сегментацию иероглифов в тексте для определения границ каждого слова и частиц в тексте. Программа, позволившая выполнить сегментацию носит название SegmentAnt.

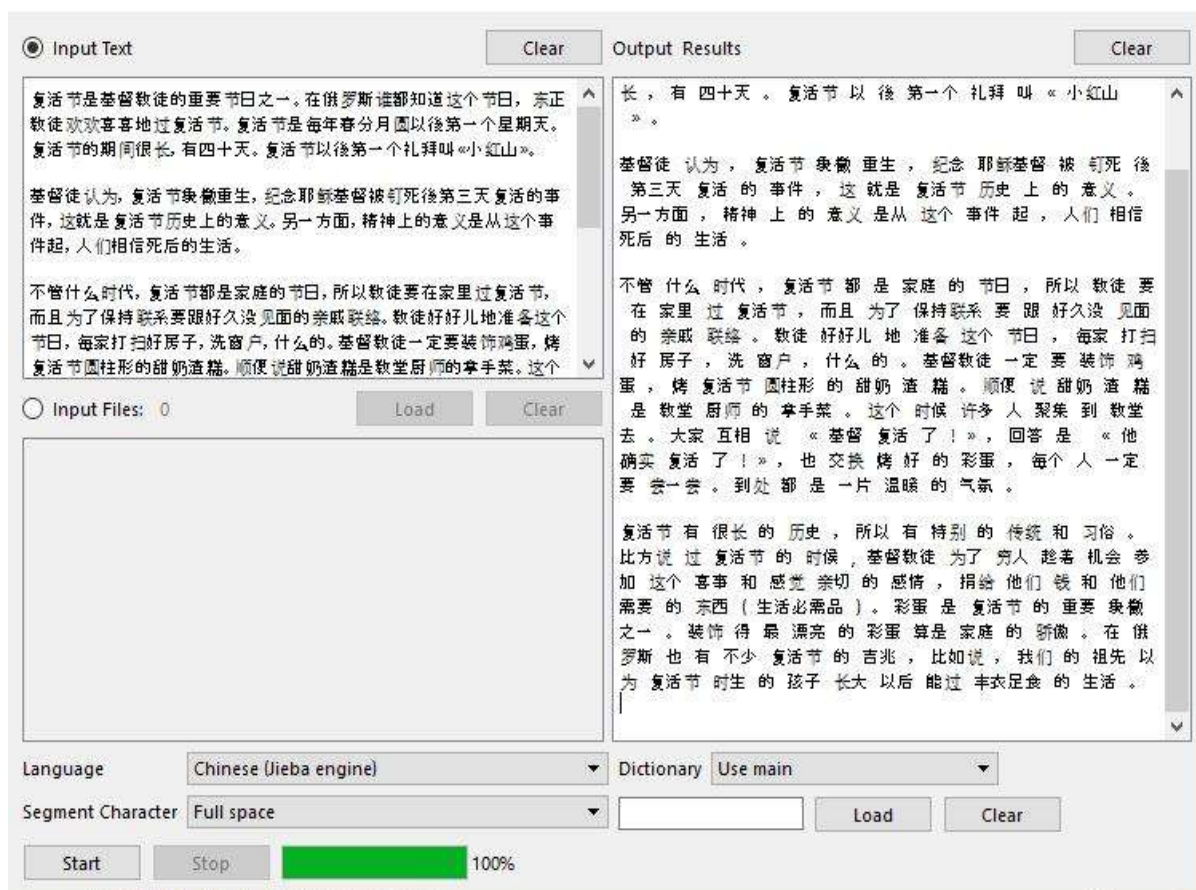


Рисунок 2.1 – Сегментация текста на китайском языке

За основу был взят текст на тему «Пасха». Начальный перевод, которому предстоит обработка будет предоставлен ниже:

*«Пасха - один из самых важных христианских праздников. Все знают этот фестиваль в России, и православные христиане празднуют Пасху с радостью. Пасха - первое воскресенье после равноденствия каждый год. Пас-*

*хальный период длительный, сорок дней. Первая неделя после Пасхи называется «Маленькая Красная Гора».*

*Христиане верят, что Пасха символизирует возрождение и воздает память о воскресении Иисуса Христа на третий день после его смерти. Это значение истории Пасхи. С другой стороны, духовный смысл состоит в том, что из этого инцидента люди верят в жизнь после смерти.*

*Независимо от возраста, Пасха - это праздник для всей семьи, поэтому христиане должны иметь Пасху дома. Чтобы поддерживать контакт, они должны общаться с родственниками, которые долго не виделись. Последователи хорошо подготовили этот фестиваль, и каждый дом убирал дом, промывал окна и что. Христиане должны украшать яйца и испечь пасхальные цилиндрические сладкие пирожные. Кстати, сладкий молочный пирог - особенное блюдо для шеф-поваров церкви. На этот раз многие люди собрались в церковь. Все говорили друг другу: «Христос воскрешен!», Ответ «Он действительно воскрес!» И обменивает испеченные яйца. Каждый должен попробовать. Во всем мире царит теплая атмосфера.*

*Пасха имеет долгую историю, поэтому существуют особые традиции и обычаи. Например, во времена Пасхи христиане воспользовались возможностью для бедных участвовать в этом счастливом событии и чувствовали ласковые чувства, давали им деньги и то, что им нужно (предметы первой необходимости). Пасхальные яйца являются одним из важных символов Пасхи. Самым красиво украшенным яйцом является гордость семьи. В России также есть много признаков Пасхи. Например, наши предки думали, что дети, родившиеся на Пасху, могут вырасти, чтобы одеться и хорошо питаться.»*

Исходя из перевода, мы можем сделать вывод, что тексту требуется доработка в согласовании и осмыслении предложений. В дальнейшем мы продолжим обрабатывать данный текст, используя выбранные нами инструменты для дальнейшего улучшения качества и скорости перевода

Программ с подобным функционалом присутствует в интернете неисчислимое количество. Как коммерческих версий, так и в свободном доступе. Для альтернативного обзора будет предоставлена ещё одна программа, однако, работающая исключительно онлайн (см. рисунок 2.2).

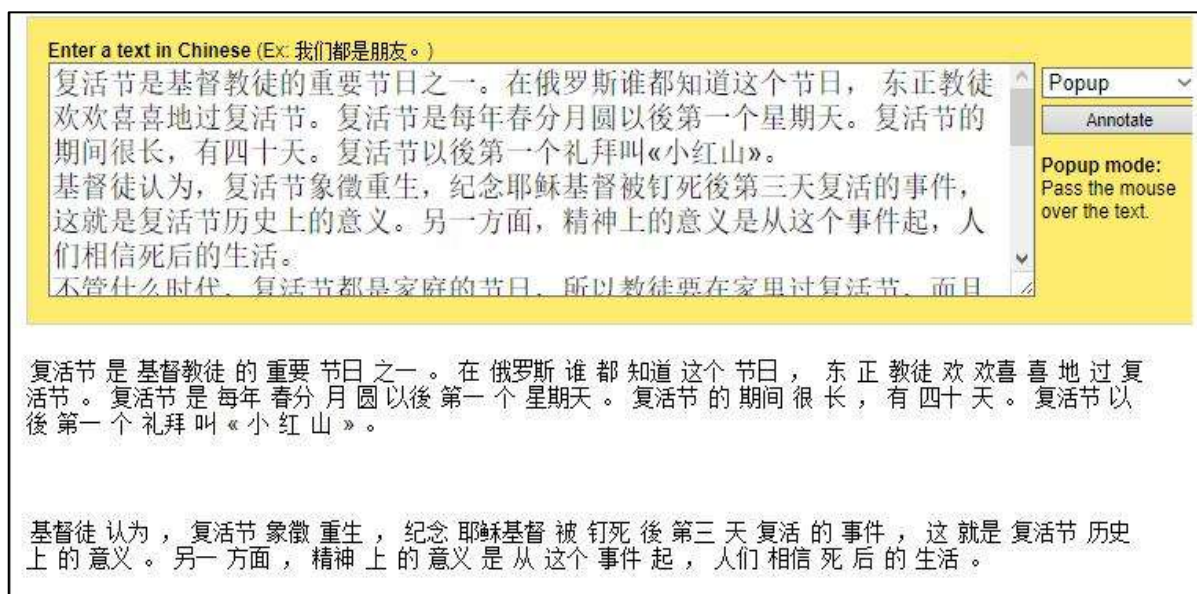


Рисунок 2.2 – Онлайн сегментация текста на китайском языке

На рисунке 2.2 изображена сегментация текста взятого за основу исследования. В поле для ввода текста был введён текст, результат его сегментации представлен ниже в рисунке. Данный инструмент является показательной альтернативой инструментария с похожим функционалом и предоставлен в свободном доступе на сайте [www.chinese-tools.com](http://www.chinese-tools.com)

### 2.1.2 Функция поиска и замены элементов в тексте

Для дальнейшего описания решений следует отметить ещё один инструмент выполняющий полезную функцию в китайских массивах текстов. Поиск и замена элементов текста представляет собой функцию, предназначенную для изменения, обобщения информации в больших массивах текстов. Также, поиск и замена элементов в текстах на китайском языке позволит повысить специфику и опыт пользователя в обработке китайских текстов. Итак, представим пример данной функции (см. рисунок 2.3).



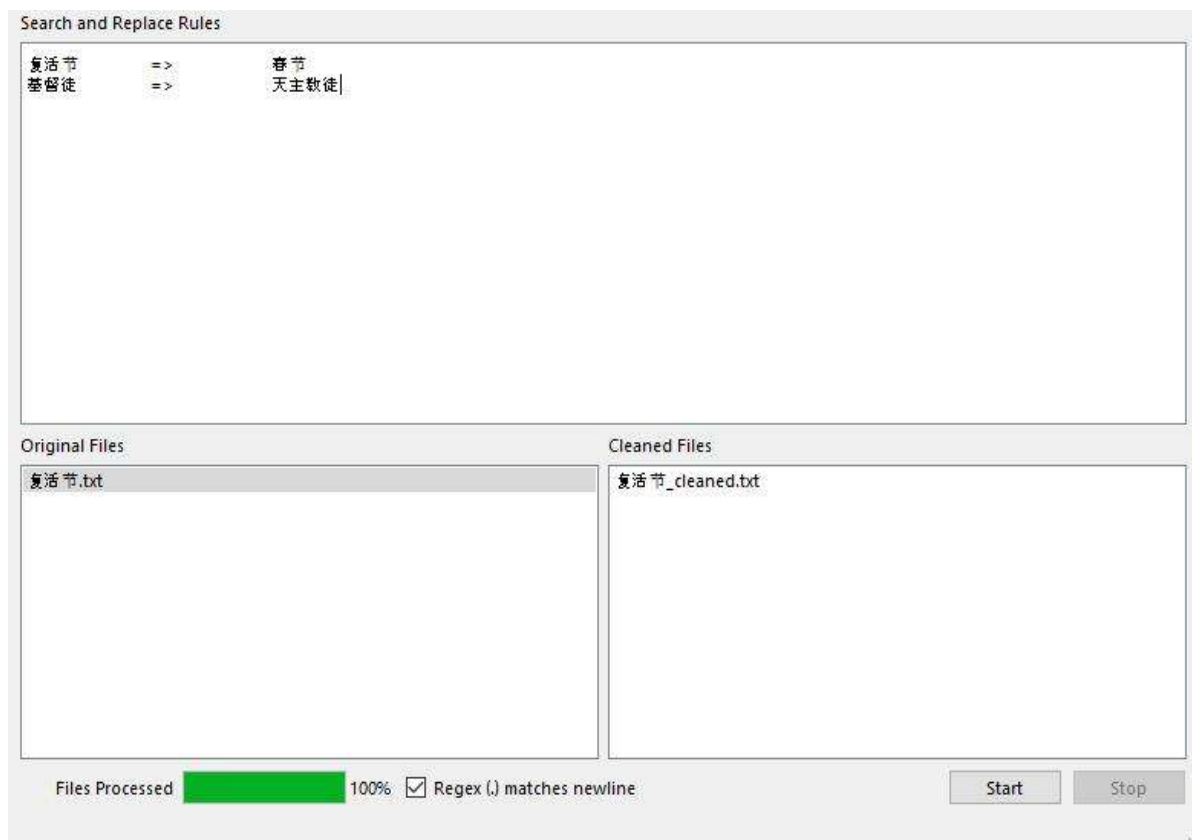


Рисунок 2.3 – Процесс заполнения правил для замены элементов в тексте

На рисунке 2.3 изображен процесс написания правила замены элементов:

- 复活节 fùhuójíe => 春节 chūnjié (Пасха => Новый год)
- 基督徒 jīdūtú => 天主教徒 tiānzhǔjiàotú (Христианин => Католик)

Оба этих правила заменяют иероглифы в тексте (复活节.txt), оригинальный текст, который пользователь хочет редактировать загружает в окно «Originals Files». После применения правил в предлагаемый текст, программа SarAnt создаёт отдельный файл (复活节\_cleaned.txt), в котором значения изменённых элементов уже применены и видны в тексте. Для просмотра оригинального и преобразованного текста обратитесь в приложение данной работы.

### 2.1.3 Поиск счётных слов для существительных

В китайском языке счётные слова являются составляющей для обозначения количества дискретных предметов. Счётные слова в китайском языке имеют свою классификацию, образуя соответствующий класс слов. В китай-

ском языке насчитывается более 100 счётных слов имеющие соответствующий класс, к которому они применяются. Подобное количество счётных слов, а также их классовое применение может запутать китаеведа в их использовании. Нами был найден интернет источник, который позволит решить проблему правильного выбора счётного слова для определённого существительного на морфологическом уровне. Для предоставления решения был зафиксирован результат (см. рисунок 2.4).

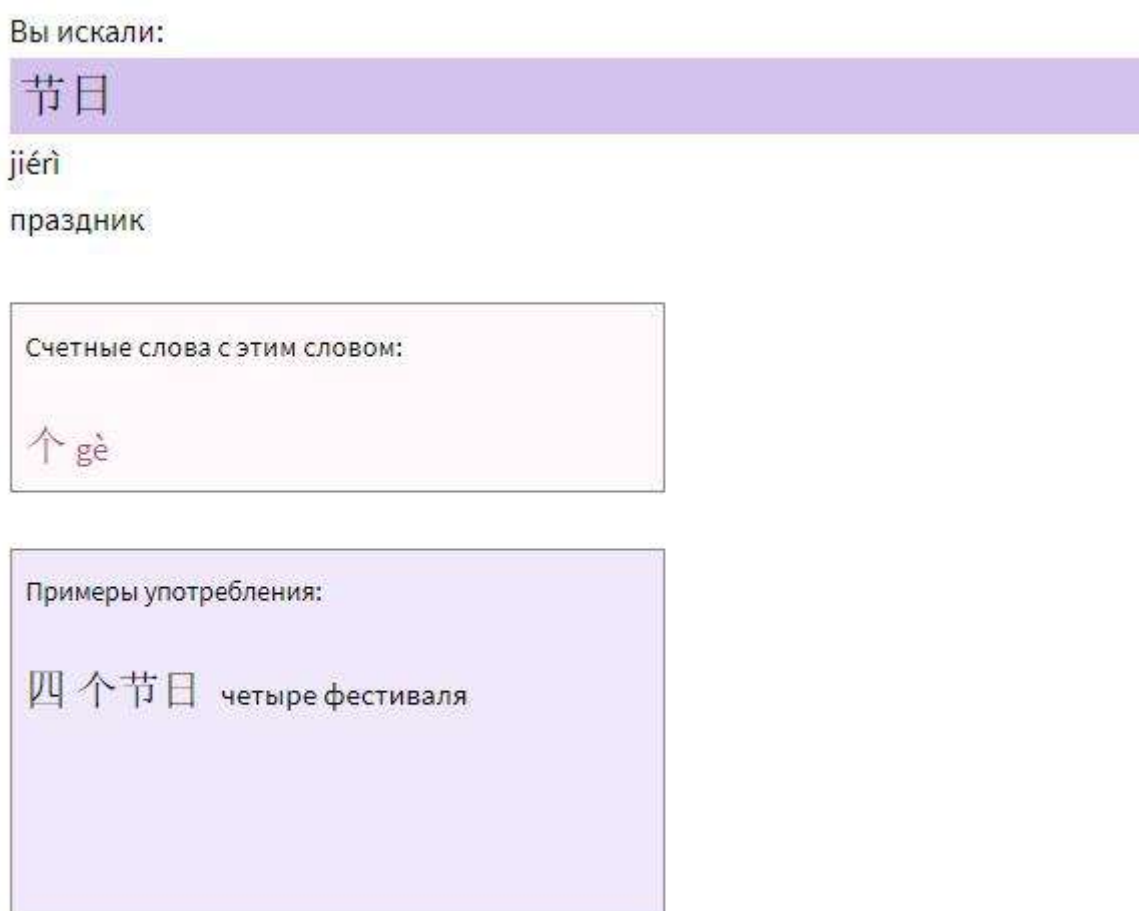


Рисунок 2.4 – Результат нахождения счётного слова

#### 2.1.4 Числительные в китайском языке

Китайский язык имеет собственную систему записи чисел. Она кардинально отличается от арабской системы написания, которая хорошо нам известна. Не смотря на повсеместное использование арабской системы числительных в китайском языке, всё же китайская система числительных приме-

няется также хорошо. Данный факт вызывает одно затруднение – преобразование китайских чисел в арабские, и наоборот. Для правильного перевода чисел из одной системы в другую требуется следовать определённым правилам, с помощью которых можно будет перевести числа на предпочитаемую систему. Но данный метод является медленным, и ошибочным при невнимательности и малом опыте. Для автоматизации, ускорения и повышения точности нами были найдены ресурсы в виде программ, которые помогают перевести числа из одной системы числительных в другую.

На рисунке 2.5 изображён результат конвертации написанных чисел из китайской системы исчисления в арабскую. Числа, продемонстрированные на скриншоте, были заимствованы из текста, с которым можно ознакомиться в приложении работы. Подобные инструменты позволят китаеведам быстро и безошибочно произвести процедуру расчёта своих данных.



Рисунок 2.5 – Конвертация китайских чисел в арабские

На рисунке 2.6 был продемонстрирован обратный процесс конвертации, однако в данном случае было применено альтернативное программное обеспечение с соответствующим функционалом. Оба инструмента являются онлайн-источниками.

Chinese/English Number(s)	Results:
18000 33 12 8843 6300 7.24 0.83	18000 = 一万八千 33 = 三十三 12 = 十二 8843 = 八千八百四十三 6300 = 六千三百 7.24 = 七点二四 0.83 = 点八三
<input type="button" value="Convert!"/> <input type="button" value="Clear"/>	
Output English as: <input type="text" value="Arabic Numbers"/> Output Chinese as: <input type="text" value="Simplified"/>	

Рисунок 2.6 – Процесс конвертации арабских чисел в китайские

## 2.2 Программное обеспечение, решающее проблемы на синтаксическом уровне китайского языка

Проблема частей речи в китайском языке – сложный вопрос. До определённого момента в истории китайские лингвисты не рассматривали вопрос синтаксиса. В китайском языке отсутствует стабильность в определении частей речи. Данный уровень китайского языка не определён как устойчивый уровень для обработки программными инструментами. В данном параграфе будет представлен инструмент, позволяющий выполнить разметку частей речи в тексте, в компьютерной лингвистике данный процесс называют «Тегирование».

Тегирование – классификация словоформ путём присвоения тегов, т.е. специальных пометок.

Часть текста «Пасха» будет взята за основу в данном испытании.

«复活节是基督教徒的重要节日之一。在俄罗斯谁都知道这个节日，东正教徒欢欢喜喜地过复活节。»

Предложения, написанные выше являются начальным материалом, перед процедурой тегирования.

На рисунке 2.7 изображен процесс маркировки предложений по частям речи. Функцию тегирования данных предложений нам помогла программа «Stanford Named Entity Recognizer». Программа с возможностью выполнения

работы онлайн. Для работы на компьютере требуются дополнительные углубленные знания языков программирования Python и Java.



Рисунок 2.7 – Процесс тегирования предложений

Для правильной расшифровки аббревиатуры наименований частей речи, изображенных над словами предоставлен лист с соответствующими значениями в приложении данной работы. Полный отчет тегирования текста «Пасха» на части речи находится в приложении работы.

### **2.3 Программное обеспечение, решающее проблемы на фонетическом уровне китайского языка**

Ни для кого не секрет, что основной фонетической особенностью китайского языка является тональность. Тональность в китайском языке напрямую связана со смысла-выражающим аспектом при общении. Обращая внимание на письменный аспект китайского языка, пользователю следует учить не только иероглифику, транскрипцию и её значение, но, а также тональность каждого из иероглифов. В соответствующем порядке транскрипция китайского языка (Система романизации «пиньинь» 汉语拼音, Hànyǔ pīnyīn) играет не малую роль при изучении и произношении китайской лексики.

Для повышения скорости и количества изучаемого материала на китайском языке, нами было изучено несколько инструментов и программ, благодаря которым мы повысим данные характеристики.

#### **2.2.1 Конвертация иероглифов в транскрипцию «Пиньинь» и установка тональности.**

В данном параграфе будут представлены несколько программ, функционал которых будет связан транскрипцией и тональностью.

Во время изучения конкретных текстов у пользователя могут появиться трудности в произношении иероглифа. Причины могут быть разными: Ма-

ленький уровень изученных иероглифов, забыл произношение иероглифа, новый выученный иероглиф не закреплённый соответствующим материалом и заданиями и т.п.

Для подобных случаев мы представляем следующие инструменты, исправляющие подобные прецеденты.

На рисунке 2.8 изображена часть большого текста на китайском языке, который пройдёт процедуру конвертации транскрипции и тональности. Данный отрезок текста уже был подвергнут процедуре сегментации, однако изображён не полностью и без результата сегментации для более удобного просмотра в письменной части выпускной работы.

复活节是基督教徒的重要节日之一。在俄罗斯谁都知道这个节日，东正教徒欢欢喜喜地过复活节。复活节是每年春分月圆以後第一个星期天。复活节的期间很长，有四十天。复活节以後第一个礼拜叫«小红山»。

基督徒认为，复活节象徵重生，纪念耶稣基督被钉死後第三天复活的事件，这就是复活节历史上的意义。另一方面，精神上的意义是从这个事件起，人们相信死后的生活。

Рисунок 2.8 – Текст перед процедурой конвертации

На рисунке 2.9 можно наблюдать транскрипцию и тональность, предназначенную для чтения конкретного отрезка текста «Пасха», изображённого на рисунке 2.8.

fù huó jié shì jī dū jiào tú de zhòng yào jié rì zhī yī. zài é luó sī  
shuí dōu zhī dào zhè gè jié rì, dōng zhèng jiào tú huān huān xǐ xǐ dì  
guò fù huó jié. fù huó jié shì měi nián chūn fēn yuè yuán yǐ hòu dì  
yī gè xīng qī tiān. fù huó jié de qī jiān hěn cháng, yǒu sì shí tiān.  
fù huó jié yǐ hòu dì yī gè lǐ bài jiào « xiǎo hóng shān » .

jī dū tú rèn wéi, fù huó jié xiàng zhēng chóng shēng, jì niàn yé sū  
jī dū bèi dīng sǐ hòu dì sān tiān fù huó de shì jiàn, zhè jiù shì fù  
huó jié lì shǐ shàng de yì yì. lìng yī fāng miàn, jīng shén shàng de  
yì yì shì cóng zhè gè shì jiàn qǐ, rén men xiāng xìn sǐ hòu de shēng  
huó.

## Рисунок 2.9 – Результат конвертации текста

Данную функцию нам помогла выполнить программа, которая носит название «NJStar Chinese WP6». Данная программа представляет из себя интерфейс текстового редактора с функциональным направлением обработки китайского языка. Конвертация иероглифов в систему транскрипции «Пиньинь» лишь одна из многих функций в данной программе. Для более точного и уверенного решения фонетической проблематики китайского языка было найдено ещё программное обеспечение решающее данную задачу. Одна из найденных программ является онлайн-конвертером. Результат представлен на рисунке 2.10.



## Рисунок 2.10 – Результат конвертации текста онлайн

Мы можем наблюдать результат конвертации китайского текста в пиньинь. Функционал данного онлайн-конвертера направлен исключительно на конвертацию китайских иероглифов в пиньинь с тонами. Данный онлайн-конвертер носит название «Chinese converter».

### 2.4 Программное обеспечение, решающее проблемы иероглифики китайского языка

Как описывается в теоретической части данной работы, в данный отрезок времени, в китайском языке используется два типа письменности:

- Традиционная письменность
- Упрощённая письменность

Использование традиционной письменности можно проследить в различных сферах жизни китайского населения, например,

- Литература
- Реклама
- Живопись
- Названия объектов и др.

На основе изученного материала, можно выявить особенность, которая может стать трудностью при переводе китайских массивов текстов. Конвертация традиционных иероглифов в упрощённую систему и наоборот станет нашим исследованием в данном параграфе. Ниже будет представлена часть текста «Пасха», который пройдёт процедуру конвертации систем записи.

«复活节是基督教徒的重要节日之一。在俄罗斯谁都知道这个节日，东正教徒欢欢喜喜地过复活节。复活节是每年春分月圆以后第一个星期天。复活节的期间很长，有四十天。复活节以后第一个礼拜叫«小红山»。

基督徒认为，复活节象征重生，纪念耶稣基督被钉死後第三天复活的事件，这就是复活节历史上的意义。另一方面，精神上的意义是从这个事件起，人们相信死后的生活。»

Изначально, текст «Пасха» записан упрощённой системой записи иероглифов, которая повсеместно используется в Китае. Представим результат конвертации систем записи.

На рисунке 2.11 изображён результат конвертации части текста «Пасха». Инструмент позволивший нам выполнить данную функцию также способен выполнить обратно-пропорциональную функцию.

復活節是基督教徒的重要節日之一。在俄羅斯誰都知道這個節日，東正教徒歡歡喜喜地過復活節。復活節是每年春分月圓以後第一個星期天。復活節的期間很長，有四十天。復活節以後第一個禮拜叫«小紅山»。

基督徒認為，復活節象徵重生，紀念耶穌基督被釘死後第三天復活的事件，這就是復活節歷史上的意義。另一方面，精神上的意義是從這個事件起，人們相信死後的生活。

Рисунок 2.11 – Традиционное письмо китайского языка



Пример из результата конвертации:

复活节 (Упр.нап.) – 復活節 (Трад.нап.) - Пасха

俄罗斯 (Упр.нап.) – 俄羅斯 (Трад.нап.) - Россия

Онлайн инструмент, размещённый на сайте [shibushi.ru](http://shibushi.ru) позволил нам выполнить эксперимент.

В приложении данной работы содержатся результаты альтернативных программных инструментов.

## **2.5 Лексикографические проблемы, встречающиеся при изучении китайского языка**

Использование словарей в китайском языке является базисом в поиске значений иероглифов. Словари компьютерные и словари печатного издания имеют свои недостатки и преимущества. Однако, идя в ногу со временем, специалисты в большей степени используют электронный тип словарей. Данный факт обусловлен тем, что электронные словари способны быстрее выдать результат перевода, также способны регулярно и быстро увеличивать базу данных, на которой основан вокабуляр словаря. Данная отрасль китайского языка массивна и имеет избыток информации, подлежащей фильтрации. Но мы представим одну из актуальных функций, способную решить проблему поиска новых слов и перевода терминологии.

### **2.5.1 Электронный словарь, как источник поиска неологизмов и терминологии**

За основу данного исследования будут взяты два словаря:

- bkrs
- zhonga

Данные словари помогут нам найти перевод новых слов, которые пользуются актуальностью в интернет-сленге. Также, мы затронем и другие области развития и жизни в Китае используя предложенные словари для поиска новых слов. Материалом исследования послужит материал статьи Ю. А. Куприяновой.

单独二孩 [dāndú èr hái] – постановление правительства КНР, разрешившее семье, в которой хотя бы один из супругов является единственным ребенком

в семье, иметь второго ребенка. Данный указ был принят в конце 2013 года.

装嫩 [zhuāng nèn] – «молодиться» (вести себя инфантильно, стремиться выглядеть моложе своих лет). Манера одеваться и вести себя в обществе отлично от основной массы свойственна молодым людям в силу возраста и всегда отражает дух времени. Субкультуры ежегодно генерируют немалое число новых слов. Некоторые современные китайки старше 30 лет стремятся выглядеть, говорить и вести себя так, словно вчера закончили школу [Ю. А. Куприянова 2016, с.9].

Y一代 [Y yīdài] – «поколением И». Так называют современную молодёжь в Китае за чрезмерную зависимость от Интернета.

Подобные сленг-выражения, а также социально-политические термины регулярно обогащают базу электронных словарей. Предоставленные примеры были апробированы через электронные словари и были сделаны выводы, регулярном обновлении баз данных слов, на которых они сформированы. В приложении данной работы предоставлены доказательства обработки новых слов в виде рисунков из использованных словарей.

## **2.6 Внеязыковые особенности обработки китайского языка**

Затронув и описав трудности, встречающиеся при обработке китайского языка на различных грамматических уровнях, мы предоставим инструменты, в большей степени, связанные с обработкой китайского языка, чем с лингвистической наукой. Также, данный параграф ознакомит каждого специалиста китаеведа с трудностями, с которыми он может встретиться, изучая китайскую лексику вне лингвистической и компьютерных наук, при этом испытать трудности, которые в дальнейшем будут препятствовать быстрому и качественному переводу. Представленные инструменты помогут специалисту в организационных моментах при переводе.

## 2.6.1 Оптическое распознавание текстов

Распознавание рукописных, печатного, а также фотографических текстов данных и перевод их в текстовые данные является актуальной темой для исследования. Распознавание латинских символов на нынешнем уровне развития подобного функционала гарантирует 99% точности распознавания, однако, иероглифам китайского языка уделено меньше времени на исследования, поэтому процент точности распознавания меньше. Подобное программное обеспечение позволит китаеведу расширить свой инструментарий и приступить к переводу текста в кратчайшие сроки.

На рисунке 2.12 изображена фотография печатного текста. Ссылаясь на вышеизложенные слова, данный текст, для обработки на компьютере, будет преобразован в текстовые данные.

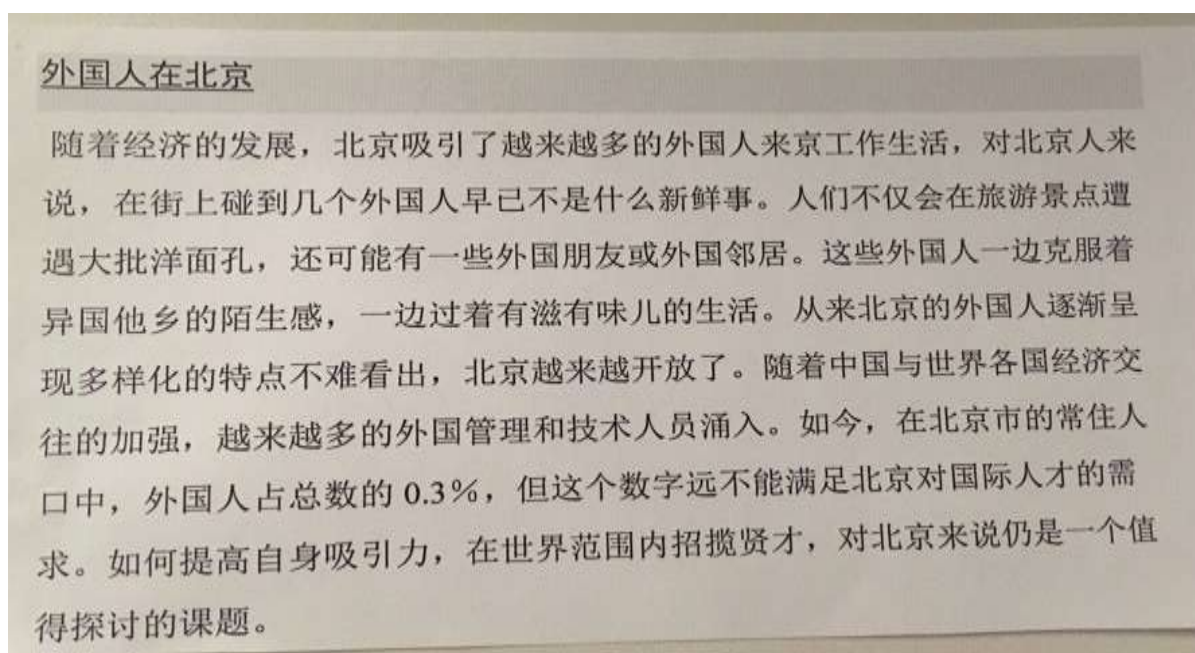


Рисунок 2.12 – Фото печатного текста перед распознаванием

С данной трудностью мы справились благодаря найденным программам, одна из которых носит название «ABBYY FineReader 14».

На рисунке 2.13 можно увидеть процесс оптического распознавания текста. Поле слева показывает оригинальную фотографию с изображённым на нём печатным текстом. Поле справа демонстрирует результат распознавания, преобразованный в электронный текст. В дальнейшем, после окончательного

процесса распознавания, окончательный результат текстовых данных можно увидеть в приложении данного исследования.

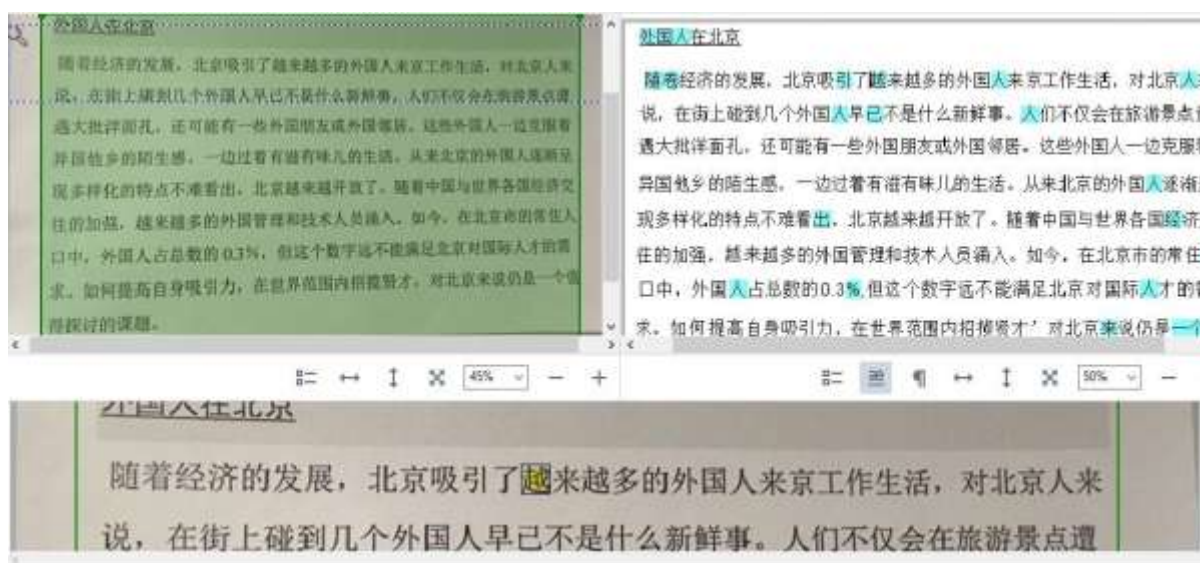


Рисунок 2.13 – Процесс распознавания печатного текста

На основе данного эксперимента можно сделать вывод, что инструмент с функцией оптического распознавания текстов будет отличным помощником в обработке языка. Благодаря данной функции китаеведу не придётся переписывать текст с печатного издания на компьютер. Подобный процесс может занять большое количество времени при обработке большого массива текстов.

## 2.6.2 Китайские единицы измерения и трудности их преобразования

Ссылаясь на теоретическую составляющую данной трудности написанную в первой главе исследования, нами были приняты соответствующие меры. Был найден инструмент, позволяющий конвертировать китайские единицы измерения в отечественную систему мер и в другие. Китайский язык, как язык, который владеет собственной системой единиц измерения для русскоязычного китаеведа способен поставить специалиста в тупик, при конвертации единиц измерения веса, длины и т.п.

Поиск решения на данную трудность обуславливается тем, что при обработке китайских массивов текста, китаеведу могут встретиться тексты с содержанием китайских систем мер (рис. 2.14).

UNITS OF LENGTH			
Value :	<input type="text" value="163"/>	From <input type="text" value="Li"/>	To <input type="text" value="Kilometers"/>
			<input type="button" value="Convertir"/>
<b>163 li = 81.5 kilometers</b> <b>一百六十三市里 = 81.5 kilometers</b>			
UNITS OF AERA			
Value :	<input type="text" value="84"/>	From <input type="text" value="Square Kilome"/>	To <input type="text" value="Qing"/>
			<input type="button" value="Convertir"/>
<b>84 square kilometers = 1260 qing</b> <b>84 square kilometers = 一千二百六十市顷</b>			
UNITS OF WEIGHT			
Value :	<input type="text" value="45"/>	From <input type="text" value="Dan 担"/>	To <input type="text" value="Kilograms"/>
			<input type="button" value="Convertir"/>
<b>45 dan = 2250 kilograms</b> <b>四十五市担 = 2250 kilograms</b>			
UNITS OF CAPACITY			
Value :	<input type="text" value="91"/>	From <input type="text" value="Liters"/>	To <input type="text" value="Sheng"/>
			<input type="button" value="Convertir"/>
<b>91 liters = 91 sheng</b> <b>91 liters = 九十一升</b>			

Рисунок 2.14 – Инструмент преобразования систем мер

Ссылаясь на рисунок 2.14 можно увидеть расширенный спектр доступных преобразований:

- Единицы измерения длины
- Единицы измерения площади
- Единицы измерения веса
- Единицы измерения ёмкости

## Выводы по главе 2

Научный прогресс, неминуемо стремящийся вперед с каждым годом открывает для нас новые возможности в различных сферах новых технологий. Во второй главе данной работы мы показали практические решения трудностей, которые были выявлены нами при написании продемонстрированной работы. Трудности китайского языка были выявлены на таких уровнях, как:

- Морфологический
- Синтаксический

- Фонетический
- Лексикографический

Также в данной главе была затронута иероглифика китайского языка и внеязыковые трудности обработки китайского языка. На каждом из этих уровней были найдены решения и предоставлены аргументы их решения в виде изображений и описания проведённого анализа. Альтернативное программное обеспечение и результаты обработки материала, приведённого на некоторых из уровней можно будет изучить в приложении данной работы.

## ЗАКЛЮЧЕНИЕ

Подводя итог нашего исследования нами была достигнута цель, поставленная в начале исследования. Инструментарий представленный в работе и в приложении преодолел трудности, поставленные нами в исследовании. Данные инструменты позволят китаеведам выполнить перевод более качественно с минимальной затратой времени. Были затронуты как лингвистические, так и программные особенности при обработке китайского языка. Задачи, поставленные в нашей работе, были выполнены поэтапно и соответствующему порядку. Написав теоретическую и практическую часть исследования можно проследить логику написания, заключающуюся в поэтапной формулировке проблематики, и в дальнейшем её практическом решении. В интернете присутствует неисчислимо количество программ с подобным функционалом. Мы постарались выявить специфичные для китайского языка. Результаты апробации материалов на китайском языке с помощью представленного в данной работе соответствует нашим ожиданиям. Каждая из программ справилась со своей задачей. Одной из основных причин написания данной работы заключалось в изучении китайских особенностей китайского языка, как проблем, с которыми может встретиться китаевед в практическом применении знаний. Соответственно, на стыке двух научных направлений – теоретической и прикладной лингвистик мы постарались создать работу, которая обогатит багаж знаний о программном инструментарии и обработке массивов текстов на китайском языке.

## БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Алексахин, А.Н. Современная политика КНР в отношении иероглифической и буквенной письменности / А.Н. Алексахин // Вестник МГИМО Университета. – 2011. – №3. – С. 243–251.
2. Юй, Чуйцяо. Статистический метод извлечения терминов из китайских текстов без сегментации фраз / Юй Чуйцяо, И.А Бессмертный, Ма Пенюй // Научно-технический вестник информационных технологий, механики и оптики. – 2016. – №6. – С. 1096–1102.
3. Ван, Ляо-и. Основы китайской грамматики [Текст]; пер. с кит. яз. Г. Н. Райской; под ред. А. А. Драгунова и Чжоу Сун-Юаня. – М.: Издательство иностранной литературы, 1954. – 262 с.
4. Головин, Б. Н. Введение в языкознание / Б.Н. Головин // М: «Высшая школа», 1966. – 329 с.
5. Горелов, В.И. Союзы в сложном предложении современного китайского литературного языка [Текст] / В.И. Горелов. – М., 1963. – 212 с.
6. Горелов, В.И. Теоретическая грамматика китайского языка [Текст]: учебное пособие для студентов пед. институтов по спец. «Иностр. яз.» / В.И. Горелов. – М.: Просвещение, 1989. – 318 с.
7. Драгунов, А. А. Исследования по грамматике современного китайского языка [Текст] / А.А. Драгунов. – М.; Л.: Изд-во АН СССР, 1952. – 231 с.
8. Иванов, А. И. Грамматика современного китайского языка [Текст] / А.И. Иванов, Е.Д. Поливанов. – Изд. 3-е, стереотипное. – М.: Едиториал УРСС, 2003. – 304 с. – (Лингвистическое наследие XX века.)
9. Ильницкая, К.А. Проблемы преподавания фонетики китайского языка русским студентам / К.А. Ильницкая // Университетские чтения. – Пятигорск: Пятигорский государственный лингвистический университет, 2015. – С. 79–83.
10. Ким, В.С. Словари китайского языка и современность / В.С. Ким, Б.Г. Фаткулин // Вестник ЧелГУ. – 2011. – №21. – С. 110 – 115.



11. Кирюхина, Л.В. Грамматические исследования в китайском языкознании // Вестник ИГЛУ. – 2014. – №1 – С. 204 – 208.
12. Коротков, Н.Н. Основные особенности морфологического строя китайского языка [Текст]: грамматическая природа слова / Н.Н. Горелов. АН СССР. Ин-т языкознания. – Москва: Наука, 1968. – 400 с.
13. Кошкин, А.П. Некоторые вопросы синтаксиса предлогов субъектно-объектных отношений китайского языка / А.П. Кошкин // ФГБОУ ВПО «Амурский гуманитарно-педагогический государственный университет» – 2013. – С. 106 – 118.
14. Кубарич, А.М. Ударение в китайском языке: экспериментально-фонетическое исследование / А.М. Кубарич // Вестник КемГУ. – 2015. – №4. – С. 110 – 114.
15. Куприянова, Ю.А., Гун Мин, Социально-экономические факторы, способствующие появлению новых слов в современном китайском языке / Ю.А. Куприянова, Гун Мин // Вестник СПбГУ. Сер. 13. Востоковедение. Африканистика. – 2016. – №1. – С. 5 – 18.
16. Курдюмов, В.А. Курс китайского языка. Теоретическая грамматика / В.А. Курдюмов. – Изд. 2-е, стереот. – М.: Цитадель – трейд: Вече, 2006. – 576 с.
17. Лазарева, О.В. Квантитативность в русском и китайском языках: имя числительное / О.В. Лазарева // Вестник РУДН. Серия: Теория языка. Семиотика. Семантика. – 2015. – №4. – С. 126 – 130.
18. Люй, Шусян. Очерк грамматики китайского языка / Люй Шусян – М.: Наука, 1961. – Том 2. – 285 с.
19. Найденова, К.А. Машинное обучение в задачах обработки естественного языка: обзор современного состояния исследований / К.А. Найденова, О.А. Невзорова // Учен. зап. Казан. ун-та. Сер. Физ.-матем. науки. – 2008. – №4. – С. 6–24.

20. Осьминина, Е. К. Сегментирование текста в системе машинного перевода / Е.К. Осьминина // Вестник ВГТУ, 2011. – №6. – С. 160–163.
21. Очиров, О.Р. Терминология современного китайского языка / О.Р. Очиров // Учёные записки ЗабГУ. Серия: Филология, история, востоковедение, 2009. – №3. – С. 236 – 238.
22. Селиванов, Ю.П. Вычислительная техника. Терминология: Справочное пособие. Выпуск 1 / Ю.П. Селиванов. – М.: Издательство стандартов, – 1989. – 168 с.
23. Симатова, С.А. Частицы как способ выражения модальности в китайском языке [Текст] / С.А. Симатова // – Вестник Санкт-петербургского университета, сер.13, 2014. – вып 14. – С. 23–26.
24. Спешнев, Н.А Фонетика китайского языка. – СПб: Издательство Ленинградского университета, 1980. – 141 с.
25. Суперанская, А.В. Общая терминология: Вопросы теории. / А.В. Суперанская, Н. В. Подольская, Н. В. Васильева // М.: Наука, – 1989. – 243 с.
26. Суперанская, А.В. Общая терминология: Терминологическая деятельность. / Н.В. Подольская, Н.В. Васильева. – Изд.3-е. – М.: Изд-во ЛКИ, – 2008. – 288 с.
27. Фаткулин, Б.Г. Использование лингвистически ориентированных модулей на языке Python для обработки больших текстовых массивов на восточных языках в целях эффективного сбора и обработки данных по отраслям востоковедческой тематики (на примере nltk) / Б.Г. Фаткулин // Вестник ЮУрГУ. Серия: Лингвистика. – 2015. – №1. – С. 72–75.
28. Фаткулин Б.Г., Прикладная лингвистика и обработка текстов на восточных языках: современные перспективы / Б.Г.Фаткулин // Вестник ЮУрГУ. Серия: Лингвистика. – 2014. – №3. – С. 15 – 18.
29. Шахматов, А.А. Синтаксис русского языка / А.А. Шахматов. – М.: Эдиториал УРСС, 2001. – 624 с.

30. Штелинг, Д.А. Целенаправленность речи и категория наклонения / Д.А. Штелинг // Русский язык за рубежом. – 1973. – № 3. – С. 64 – 70.
31. Юй, Чуцяо. Автоматический синтаксический анализ китайских предложений при ограниченном словаре / Юй Чуцяо, И.А. Бессмертный // Программные продукты и системы. – 2017. – №1 – С. 138 – 142.
32. Янкивер, С. Б. Предлоги в китайском языке [Текст]: Автореф. дисс.. канд филол наук / С.Б. Янкивер // АН СССР. Ин-т народов Азии. – Москва: [б. и], 1964. – 16 с.
33. Anthony Lawrence software [Электронный ресурс]. – Режим доступа: <http://www.laurenceanthony.net/software.html>. – Заглавие с экрана. – (Дата обращения: 6.02.2018).
34. Chinese – Tools [Электронный ресурс]. – Режим доступа: <https://www.chinese-tools.com>. – Заглавие с экрана. – (Дата обращения: 09.03.2018).
35. Shibushi [Электронный ресурс]. – Режим доступа: <http://www.shibushi.ru/tools>. – Заглавие с экрана. – (Дата обращения: 18.03.2018).
36. ChineseTools [Электронный ресурс]. – Режим доступа: <https://www.chinesetools.eu/tools/unit-converter>. – Заглавие с экрана. – (Дата обращение: 29.03.2018).
37. On-line Chinese Tools [Электронный ресурс]. – Режим доступа: <http://www.mandarintools.com>. – Заглавие с экрана. – (Дата обращение: 30.03.2018).
38. Chinese Converter [Электронный ресурс]. – Режим доступа: <https://www.chinesconverter.com/en>. – Заглавие с экрана. – (Дата обращения: 03.04.2018)
39. Toshuo [Электронный ресурс]. – Режим доступа: <https://toshuo.com>. – заглавие с экрана. – (Дата обращения: 11.04.2018).

40. The Stanford Natural Language Processing Group [Электронный ресурс]. – Режим доступа: <https://nlp.stanford.edu>. – заглавие с экрана. – (Дата обращения: 21.04.2018).

41. Alphabetical list of part-of-speech [Электронный ресурс]. – Режим доступа: [https://www.ling.upenn.edu/courses/Fall\\_2003/ling001/penn\\_treebank\\_pos.html](https://www.ling.upenn.edu/courses/Fall_2003/ling001/penn_treebank_pos.html). – заглавие с экрана. – (Дата обращения: 05.05.2018)

42. Большой китайско-русский словарь [Электронный ресурс]. – Режим доступа: <https://bkrs.info>. – Заглавие с экрана. – (Дата обращения: 24.05.2018).

## ПРИЛОЖЕНИЕ

### Результат апробации текста «Пасха»

#### «Праздник пасхи

Пасха является одним из важнейших православных праздников. В России все знают этот праздник, и православные необычайно радостно встречают праздник Пасхи. Пасха каждый год празднуется в разные числа месяца, но обязательно в воскресный день. Период празднования Пасхи очень длинный – 40 дней. Первая неделя после Пасхи называется «Красная Горка».

Православные считают, что праздник Пасхи символизирует возрождение, отмечается воскрешение Иисуса Христа на третий день после его погребения, в этом и заключается исторический смысл Пасхи. Духовный же смысл Пасхи заключается в том, что после этого события люди стали верить в жизнь после смерти.

В любое время, Пасха всегда считалась семейным праздником. Отмечали ее в кругу семьи, а также ходили в гости к родственникам, которых давно не навещали. Верующие также тщательно готовятся к Пасхе: убирают квартиру, моют окна и т.д. Обязательно красят яйца, пекут куличи и пасхи. Кстати говоря, кулич является самым традиционным блюдом Пасхи. В ночь Христова воскресения в храмы стекается много людей. Все православные приветствуют друг друга словами “Христос воскрес!” и отвечают “Воистину воскрес!”, а также обмениваются выпеченными куличами и пасхами, каждый обязательно должен попробовать. Везде атмосфера тепла и уюта.

Празднику этому много сотен лет, он имеет свои особые традиции и обычаи. Например, в дни Пасхи у христиан было принято раздавать бедным деньги и необходимые вещи (предметы первой необходимости), чтобы и они могли стать в этот светлый праздник участниками всеобщей радости. Одним

из главных символов праздника является яйцо. Красиво раскрашенные яйца считались гордостью семьи. В России есть немало хороших примет, например, считалось, что тот, кто родился в пасхальную неделю, будет жить в полном достатке.»

Исходя из прочитанного перевода текста, можно сделать вывод, что представленное программное обеспечение помогло выполнить более качественный перевод.

### **Результат работы программы SarAnt на материале текста «Пасха»**

В данном приложении будет представлен отрезок текста «Пасха» прошедший апробацию по замене иероглифов в тексте

«复活节是基督教徒的重要节日之一。在俄罗斯谁都知道这个节日，东正教徒欢欢喜喜地过复活节。复活节是每年春分月圆以后第一个星期天。复活节的期间很长，有四十天。复活节以后第一个礼拜叫«小红山»。»

Далее будет представлен текст с применением правил, указанных практической части работы

«春节是天主教徒的重要节日之一。在俄罗斯谁都知道这个节日，东正教徒欢欢喜喜地过 春节。 春节是每年春分月圆以后第一个星期天。 春节的期间很长，有四十天。 春节以后第一个礼拜叫«小红山»。»

В тексте курсивом указаны слова, подчеркнутые курсивом, которые были заменены.

### **Таблица аббревиатур частей речи**

Tag	Description
CC	Coordinating conjunction
CD	Cardinal number
DT	Determiner

EX	Existential <i>there</i>
FW	Foreign word
IN	Preposition or subordinating conjunction
JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
LS	List item marker
MD	Modal
NN	Noun, singular or mass
NNS	Noun, plural
NNP	Proper noun, singular
NNPS	Proper noun, plural
PDT	Predeterminer
POS	Possessive ending
PRP	Personal pronoun
PRP\$	Possessive pronoun
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
RP	Particle
SYM	Symbol
TO	<i>to</i>
UH	Interjection
VB	Verb, base form
VBD	Verb, past tense
VBG	Verb, gerund or present participle
VBN	Verb, past participle

VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
WDT	Wh-determiner
WP	Wh-pronoun
WP\$	Possessive wh-pronoun
WRB	Wh-adverb

Таблица 1 – Аббревиатура частей речи, использующиеся при тегировании

### Тегирование текста «Пасха» по частям речи



Рисунок 1 – Тегирование текста «Пасха» по частям речи

### Конвертация письменности текста «Пасха»

«复活节是基督教徒的重要节日之一。在俄罗斯谁都知道这个节日，东正教徒欢欢喜喜地过复活节。复活节是每年春分月圆以后第一个星期天。复活节的期间很长，有四十天。复活节以后第一个礼拜叫«小红山»。

基督徒认为，复活节象征重生，纪念耶稣基督被钉死后第三天复活的事件，这就是复活节历史上的意义。另一方面，精神上的意义是从这个事件起，人们相信死后的生活。



不管什么时代，复活节都是家庭的节日，所以教徒要在家里过复活节，而且为了保持联系要跟好久没见面的亲戚联络。教徒好好儿地准备这个节日，每家打扫好房子，洗窗户，什么的。基督教徒一定要装饰鸡蛋，烤复活节圆柱形的甜奶渣糕。顺便说甜奶渣糕是教堂厨师的拿手菜。这个时候许多人聚集到教堂去。大家互相说 «基督复活了!», 回答是 «他确实复活了!», 也交换烤好的彩蛋，每个人一定要尝一尝。到处都是一片温暖的气氛。»

На рисунке 2 изображён альтернативный инструмент, выполняющий функцию преобразования иероглифов на примере текста «Пасха».



Рисунок 2 – Текст «Пасха», преобразованный в традиционные иероглифы

## Лексикографическая составляющая китайских словарей

перевести

单独二胎  
dāndú èr hái

родитель единственный ребёнок может заводить двоих детей (демографическая политика КНР)

Редактировать

ссылки с:  
单独二胎

пословный:

单独 dāndú один; отдельный, одионый; изолированный, сепаратный; единоличный; самостоятельный	二胎 èrhái второй ребёнок, два ребёнка
---	--

похожие:  
单独二胎

Рисунок 3 – Результат поиска термина в словаре БКРС

перевести

装嫩  
zhuāng nèn

стремиться выглядеть моложе своих лет, не по возрасту (напр. о манере говорить, одеваться и т.д.)

Редактировать

zhuāng nèn  
to act young  
to affect a youthful appearance

Рисунок 2 – Результат поиска неологизма в словаре БКРС