БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

Факультет прикладной математики и информатики

Кафедра информационных систем управления

ГРИГОРЬЕВА МАРИЯ ГРИГОРЬЕВНА

**СИСТЕМА АВТОМАТИЗАЦИИ ОБУЧЕНИЯ ГРАММАТИКЕ АНГЛИЙСКОГО ЯЗЫКА**

Курсовой проект

студентки 4 курса 2 группы

|  |  |
| --- | --- |
|  | Руководитель  Совпель Игорь Васильевич |

Минск 2017

**Аннотация**

В курсовом проекте рассмотрена проблема автоматизации обучения английскому языку посредством формализации грамматики. На основе анализа литературы определены основные проблемы реализации распознавания и компьютерного представления словоформ. Разработана программа, определяющая времена предложений английского языка.

**Анатацыя**

У працы разгледжана праблема аўтаматызацыі навучання англійскай мове праз фармалізацыю граматыкі. На аснове аналізу літаратуры вызначаны асноўныя праблемы рэалізацыі распазнавання і камп’ютарнага прадстаўлення славаформ. Распрацавана праграма, якая вызначае час сказаў у ангійскай мове.

**Annotation**

In the course project the automated education of English grammar problem was examined. Based on the literature review identified the main problems of implementation of recognition and computer representation of word forms. The program of the English tenses determination was developed.

**Реферат**

Курсовой проект, стр., рис., табл., скриншотов, источников.

СИСТЕМА АВТОМАТИЗАЦИИ ОБУЧЕНИЯ ГРАММАТИКЕ АНГЛИЙСКОГО ЯЗЫКА

Объектом исследования является формализированное представление грамматики английского языка в компьютерной системе.

Цель работы:

1. Изучить литературу по данной теме
2. Изучить способы представления грамматических форм языка
3. Программно реализовать компьютерную обработку словоформ английского языка
4. Описать полученные результаты
5. Подготовить отчетный документ по проведенной работе

Методология проведения работы: системный подход, компьютерная лингвистика, инженерия знаний, технологии обработки естественных языков.

Область применения результатов: в разработке программного обеспечения для обучения грамматики английского языка.

**СОДЕРЖАНИЕ**

**ВВЕДЕНИЕ** …………………………………………………………………...7

**ГЛАВА 1. ВРЕМЕНА В ГРАММАТИКЕ АНГЛИЙСКОГО ЯЗЫКА**………………………………………………………………………...……..8

1.1 Введение

1.2 Настоящее время (Present)

1.3 Прошедшее время (Past)

1.4 Будущее время (Future)

1.5 Выводы

**ГЛАВА 2. ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ POS-ТЕГОВ В РАСПОЗНОВАНИИ СЛОВОФОРМ**………………………………

1.1 Общие сведения

1.2 Применение частеречной разметки для определения времени предложения

1.3 Выводы

**ГЛАВА 3. ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ РЕГУЛЯРНЫХ ВЫРАЖЕНИЙ**……………………………

**ЗАКЛЮЧЕНИЕ**…………………………………………..………………

**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**…………..……..

**ПРИЛОЖЕНИЕ А**………………………………………………………..

А.1 Постановка прикладной задачи………………………………...……20

А.2 Результаты реализации задачи на ОС Windows…………………….20

А.3 Результаты реализация задачи на ОС UNIX – Linux……………….22

А.4 Сравнительный анализ работы программ…………………………...23

# ВВЕДЕНИЕ

Тема курсового проекта посвящена проблеме автоматизации обучения английскому языку посредством формализации грамматики, а также реализации распознавания и компьютерного представления словоформ.

Несмотря на большие объемы исследований, данная тема является актуальной, так как появляются новые поколения задач в связи с процессом глобализации, новыми достижениями в сфере IT и коммуникаций. В данной работе будут затронуты темы, являющиеся одними из ключевых в компьютерной лингвистике. Представленные концепции помогают решить проблемы обработки естественного языка, такие как машинный перевод, распознавание речи, анализ, генерирование, автоматическое реферирование, аннотирование или упрощение текста.

**ГЛАВА 1. ВРЕМЕНА В ГРАММАТИКЕ АНГЛИЙСКОГО ЯЗЫКА**

* 1. Введение

В данной главе рассмотрим 12 времен грамматики современного английского языка.

* 1. Настоящее время (Present)

1. Present Simple

Время **Present Simple** обозначает действие в настоящем в широком смысле слова. Оно употребляется для обозначения обычных, регулярно повторяющихся или постоянных действий, например, когда мы говорим о чьих-либо привычках, режиме дня, расписании и т. д., т. е. Present Simple обозначает действия, которые происходят в настоящее время, но не привязаны именно к моменту речи.

Английский глагол во временной форме Present Simple почти всегда совпадает со своей начальной, то есть указанной в словаре, формой без частицы to. Лишь в 3-ем лице единственного числа к ней нужно прибавить окончание –s:

I work – he work**s**

Для того, чтобы построить **вопросительное предложение**, перед подлежащим нужно поставить вспомогательный глагол. Время Present Simple используется без него, поэтому в этом случае добавляется вспомогательный глагол do (или does в 3 л. ед. ч.).

Do you like rock?

Does he speak English?

В **отрицательных предложениях** тоже используется вспомогательный глагол do/does, но не перед подлежащим, а перед глаголом. После него прибавляется отрицательная частица not. Do/does и not часто сокращаются до don’t и doesn’t соответственно.

I do not like black coffee.

She doesn't smoke.

б) Present Continuous

Времена группы **Continuous** указывают на процесс, действие, длящееся в определенный момент в прошлом, настоящем или будущем.

Время **Present Continuous** обычно указывает на процесс, длящийся непосредственно в момент речи. На это могут указывать контекст или такие слова, как now (сейчас), at the moment (в текущий момент) и т.п.

Для того, чтобы поставить глагол в форму времени Present Continuous, требуется вспомогательный [глагол to be](https://www.native-english.ru/grammar/verb-to-be) в настоящем времени и [причастие настоящего времени](https://www.native-english.ru/grammar/participle) (Participle I) смыслового глагола.

To be в настоящем времени имеет три формы:

* am – 1 лицо, ед. ч. (I am shaving.)
* is – 3 лицо, ед. ч. (He is  reading.)
* are – 2 лицо ед. ч. и все формы мн. ч. (They are sleeping.)

Причастие настоящего времени (Participle I) можно получить, прибавив к начальной форме значимого глагола окончание -ing:

jump – jumping, live – living

В **вопросительном предложении** вспомогательный глагол выносится на место перед подлежащим, а значимый глагол остается после него:

Why are you laughing?

Are you using this dictionary?

В **отрицательных предложениях** за вспомогательным глаголом следует отрицательная частица not. Формы is и are при этом могут быть сокращены до isn’t и aren’t соответственно:

Radio is not (isn’t) working.

в) Present Perfect

Время **Present Perfect** обозначает действие, которое завершилось к настоящему моменту или завершено в период настоящего времени. Хотя [английские глаголы](https://www.native-english.ru/grammar/english-verbs) в Present Perfect обычно переводятся на русский язык в прошедшем времени, следует помнить, что в английском языке эти действия воспринимаются в настоящем времени, так как привязаны к настоящему результатом этого действия.

I have done my homework already.

We have no classes today, our teacher has fallen ill.

Время Present Perfect образуется при помощи вспомогательного [глагола to have](https://www.native-english.ru/grammar/verb-to-have) в настоящем времени и [причастия прошедшего времени](https://www.native-english.ru/grammar/participle) значимого глагола, то есть его «третьей формы».

To have в настоящем времени имеет две формы:

* has – 3 лицо, ед. ч. (He has played)
* have – 1 и 2 лицо ед.ч. и все формы мн. ч. (I have played)

Причастие второе, или причастие прошедшего времени (Participle II), можно получить, прибавив к начальной форме значимого глагола окончание -ed:

examine – examined, enjoy – enjoyed, close – closed

Однако есть также достаточно большая группа неправильных английских глаголов, которые образуют форму прошедшего времени не по общим правилам. Для них форму причастия прошедшего времени нужно запомнить.

В **вопросительном предложении** вспомогательный глагол выносится на место перед подлежащим, а значимый глагол остается после него:

Have you seen this film?

Has she come yet?

В **отрицательных предложениях** за вспомогательным глаголом следует отрицательная частица not. Формы have и has при этом могут быть сокращены до haven’t и hasn’t соответственно:

I haven’t seen him since then.

г) Present Perfect Continuous

Время **Present Perfect Continuous** указывает на действие, которое началось в прошлом, продолжалось в течение некоторого времени и либо закончилось непосредственно перед разговором или все еще продолжается в момент разговора.

I have been waiting here for 2 hours!

We have been preparing for our exam since morning.

Для того, чтобы поставить глагол в форму времени Present Perfect Continuous, требуется вспомогательный глагол to be во времени Present Perfect и причастие настоящего времени (форма V-ing) смыслового глагола.

To be во времени Present Perfect имеет две формы:

Have been – 1 и 2 лицо ед. ч. и все формы мн. ч.

Has been – 3 лицо ед. ч.

Причастие настоящего времени (Participle I) можно получить, прибавив к начальной форме значимого глагола окончание -ing:

jump – jumping, live – living

В **вопросительном предложении** вспомогательный глагол to have выносится на место перед подлежащим, а остальная часть сказуемого располагается после него:

I smell tobacco. Have you been smoking?

Has she been using my car again?

В **отрицательных предложениях** за вспомогательным глаголом to have следует отрицательная частица not:

It has not been snowing here since 1993.

* 1. Прошедшее время (Past)

1. Past Simple

Время **Past Simple** используется для обозначения действия, которое произошло в определенное время в прошлом и время совершения которого уже истекло. Для уточнения момента совершения действия в прошлом при использовании времени **Past Simple** обычно используются такие слова, как five days ago (пять дней назад), last year (в прошлом году), yesterday (вчера), in 1980 (в 1980 году) и т.п.

Для того, чтобы поставить английский глагол во время **Past Simple**, нужно использовать его «вторую форму». Для большинства глаголов она образуется прибавлением окончания -ed:

examine – examined, enjoy – enjoyed, close – closed

Однако есть также достаточно большая группа неправильных английских глаголов, которые образуют форму прошедшего времени не по общим правилам, для них форму прошедшего времени нужно просто запомнить.

We saw your dog two blocks from here.

В **вопросительном предложении** перед подлежащим нужно использовать вспомогательный глагол do в прошедшем времени – did, а после подлежащего поставить основной, значимый глагол в начальной форме:

Did you wash your hands?

Did they sign the contract?

В **отрицательных предложениях** перед глаголом нужно поставить вспомогательный глагол did и отрицательную частицу not:

We did not find our car.

I did not understand this question.

б) Past Continuous

Время **Past Continuous** указывает на процесс, длившийся в определенный момент или период в прошлом. В отличие от времени Past Simple, этот момент в прошлом должен быть назван прямо (например, yesterday at 5 o’clock, when you called, when rain started) или быть очевидным из контекста.

When you called I was taking a shower.

Charlie and me were already driving home when the engine suddenly stopped.

Для того, чтобы поставить глагол в форму Past Continuous, нужен вспомогательный глагол to be в прошедшем времени и причастие настоящего времени (Participle I) смыслового глагола.

To be в прошедшем времени имеет две формы:

was – 1 и 3 лицо ед. ч. (I was smoking. He was eating.)

were – 2 лицо ед. ч. и все формы мн. ч. (They were laughing.)

В **вопросительном предложении** вспомогательный глагол выносится на место перед подлежащим, а значимый глагол остается после него:

Were you sleeping when I called? You sounded drowsy.

В **отрицательных предложениях** за вспомогательным глаголом следует отрицательная частица not. Формы was и were при этом могут быть сокращены до wasn’t и weren’t соответственно.

In the morning our elevator was not (wasn’t) working.

в) Past Perfect

Время **Past Perfect** обозначает действие, которое завершилось до некоего момента в прошлом:

I called Jim too late, he had already left.

We had lived in Paris for 12 years before we moved to America.

Время Past Perfect образуется при помощи вспомогательного глагола to have в прошедшем времени и причастия прошедшего времени значимого глагола, то есть его «третьей формы».

To have в прошедшем времени имеет единственную форму had.

Причастие второе, или причастие прошедшего времени (Participle II), можно получить, прибавив к начальной форме значимого глагола окончание -ed:

examine – examined, enjoy – enjoyed, close – closed

Однако в английском языке есть также достаточно большая группа неправильных глаголов, которые образуют форму прошедшего времени не по общим правилам.

В **вопросительном предложении** вспомогательный глагол выносится на место перед подлежащим, а значимый глагол остается после него:

Had you brushed your teeth before you went to bed?

В **отрицательных предложениях** за вспомогательным глаголом следует отрицательная частица not. При этом они могут быть сокращены до формы hadn’t.

How did you hope to pass the exam if you had not (hadn’t) even opened the textbook?

г) Past Perfect Continuous

Время **Past Perfect Continuous** указывает на действие, которое началось в прошлом, продолжалось в течение некоторого времени и либо закончилось непосредственно перед неким моментом в прошлом или все еще не закончилось к некоему моменту в прошлом.

I had been typing this text for 2 hours and then found it on the Internet.

I had been waiting for his airplane for 2 hours when it was announced about delay.

Для того, чтобы поставить глагол в форму времени Past Perfect Continuous, требуется вспомогательный глагол to be во времени Past Perfect и причастие настоящего времени (форма V-ing) смыслового глагола.

To be во времени Past Perfect имеет единственную форму had been.

Причастие настоящего времени (Participle I) можно получить, прибавив к начальной форме значимого глагола окончание -ing:

jump – jumping, live – living

В **вопросительном предложении** вспомогательный глагол had выносится на место перед подлежащим, а остальная часть сказуемого располагается после него:

I saw many puddles. Had it been raining?

В **отрицательных предложениях** за вспомогательным глаголом had следует отрицательная частица not:

Of course he did not want to sleep! He had not been working like a horse all the week.

* 1. Будущее время (Future)

а) Future Simple

Время **Future Simple** ссылается на действие, которое совершится в неопределенном или отдаленном будущем. Простое будущее время обычно используется с обстоятельствами: tomorrow (завтра), next year (в следующем году), in five years (через пять лет), in 2035 (в 2035 году) и т.п.

Для того, чтобы поставить глагол во временную форму Future Simple, нужно использовать его начальную форму и вспомогательный глагол shall (для первого лица) или will (второе и третье лицо). В устной речи shall и will чаще всего сокращаются до формы ‘ll, которая может использоваться во всех лицах.

Примечание:

В современном английском, особенно в устной речи, will стал использоваться и в первом лице:

I will go to Shanghai next summer.

В **вопросительном предложении** вспомогательные глаголы shall или will ставятся перед подлежащим. Значимый глагол остается после подлежащего в своей начальной форме:

Shall we go to the beach?

Will your boss agree with our conditions?

В **отрицательных предложениях** за вспомогательным глаголом следует отрицательная частица not. Вместе они могут быть сокращены до формы shan’t (чаще в брит. английском) или won’t:

I shall not (shan’t) let you down.

Fred will not (won’t) agree to help us.

б) Future Continuous

Время **Future Continuous** указывает на процесс, который будет длиться в определенный момент в будущем. В отличие от времени Future Simple, этот момент в будущем должен быть назван прямо (tomorrow at 4 o’clock, when we meet) или быть очевидным из контекста.

Tomorrow this time I shall be flying to New York.

Don’t disturb me in the evening, I’ll be preparing for exam.

Для того, чтобы поставить глагол в форму **Future Continuous**, нужен вспомогательный глагол to be в будущем времени и причастие настоящего времени (Participle I) смыслового глагола.

To be в будущем времени имеет две формы:

shall be – 1 лицо. (I shall be studying. We shall be running.)

will be – 2 и 3 лицо. (They will be packing presents.)

Примечание:

Личные местоимения и вспомогательный глагол часто сокращаются до I’ll, he’ll, they’ll и т. д.

В **вопросительном предложении** shall / will выносится на место перед подлежащим, а to be и смысловой глагол остаются после него:

Why did you bring an album? Shall we be studying painting?

В **отрицательных предложениях** после shall / will ставится отрицательная частица not. Вместе они могут быть сокращены до формы shan’t (чаще в брит. английском) или won’t:

At 6 I won’t be sleeping yet.

в) Future Perfect

Время **Future Perfect** используется довольно редко, оно обозначает действие, которое закончится до определенного момента или начала другого действия в будущем или будет продолжать длиться после него.

Next year we shall have been married for 30 years.

Время Future Perfect образуется при помощи вспомогательного глагола to have в будущем времени и причастия прошедшего времени значимого глагола, то есть его «третьей формы».

To have в будущем времени имеет две формы:

shall have – 1 лицо. (We shall have covered 30 km by sunset)

will have – 2 и 3 лицо. (They will have covered 30 km by sunset)

Причастие второе, или причастие прошедшего времени (Participle II), можно получить, прибавив к начальной форме значимого глагола окончание -ed:

examine – examined, enjoy – enjoyed, close – closed

Однако в английском языке есть достаточно большая группа неправильных глаголов, которые образуют форму прошедшего времени не по общим правилам и которые необходимо запомнить.

В **вопросительном предложении** вспомогательный глагол выносится на место перед подлежащим, а значимый глагол остается после него:

Will you have read all these books by the exam time?

В **отрицательных предложениях** за вспомогательным глаголом следует отрицательная частица not. При этом они могут быть сокращены до формы shan’t (won’t):

I guess, I shall not have received your next letter before Christmas.

г) Future Perfect Continuous

Время **Future Perfect Continuous** указывает на действие, которое началось и продолжалось в течение некоторого времени до определенного момента в будущем. Это время используется очень редко, а в устной речи – практически никогда.

Next month we shall have been living together for 25 years.

Для того, чтобы поставить глагол в форму времени Future Perfect Continuous, требуется вспомогательный глагол to be во времени Future Perfect и причастие настоящего времени (форма V-ing) смыслового глагола.

To be во времени Future Perfect имеет две формы:

Shall have been – 1 лицо.

Will have been – 2 и 3 лицо.

Причастие настоящего времени (Participle I) можно получить, прибавив к начальной форме значимого глагола окончание -ing:

jump – jumping, live – living

В **вопросительном предложении** вспомогательный глагол shall/will выносится на место перед подлежащим, а остальная часть сказуемого располагается после него:

I hear, you are going to retire next month? For how many years will you have been working here by that time?

В **отрицательных предложениях** за вспомогательным глаголом shall/will следует отрицательная частица not:

By 2015 you will not have been living in this country long enough to participate in the election.[1]

* 1. Выводы

В современной грамматике английского языка существует 12 основных времен. Все они имеют похожую структуру и в какой-то степени пересекаются между собой. Закономерности можно проследить в табл. 1.

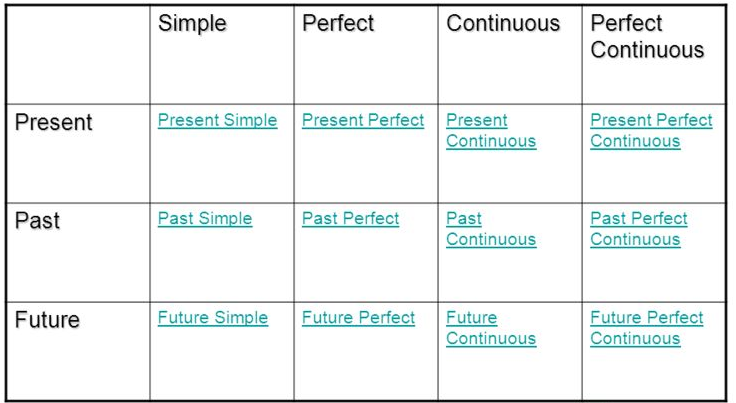


Таблица 1

**ГЛАВА 2. ОСОБЕННОСТИ ИСПОЛЬЗОВАНИЯ POS-ТЕГОВ В РАСПОЗНОВАНИИ СЛОВОФОРМ**

2.1 Общие сведения

Частеречная разметка (автоматическая морфологическая разметка, POS tagging, part-of-speech tagging) – этап автоматической обработки текста, задачей которого является определение части речи и грамматических характеристик слов в тексте (корпусе) с приписыванием им соответствующих тегов, основанных как на определении слова, так и на его контексте, т.е. на его связи со смежными и связанными словами во фразе, предложении или параграфе. POS tagging является одним из первых этапов компьютерного анализа текста. [2]

Частеречная разметка сложнее, чем просто список слов и их частей речи, потому что некоторые слова могут представлять более одной части речи в разных позициях. Это не редкость — в естественных языках (в отличие от многих искусственных языков) большой процент текстовых форм неоднозначен. Для английского языка типично выделять от 50 до 150 отдельных частей речи. [3]

Алгоритмы POS tagging делятся на две группы: основанные на правилах и вероятностные. [2]

Полученный результат после частеречной разметки представляет собой список слов и соответствующих им наборов кодов (POS тегов), которые указывают на принадлежность слова к определенной части речи. В состав кода входит: 1) основа кода (например, N… – имя существительное); 2) суффиксы, обозначающие подкласс (например, NN – имя существительное, нарицательное, ед.ч., NNP – имя существительное, нарицательное, ед.ч. с заглавной первой буквой, NNPL – язык программирования). [4]

2.2 Применение частеречной разметки для определения времени предложения

В связи с тем, что на данный момент лучшей библиотекой, разработанной для компьютерной лингвистики, является NLTK, написанная на языке программирования Python, я использовала именно его в своих исследованиях.

В библиотеке NLTK можно найти самый распространенный на данный момент POS tagger. К сожалению, основываясь на результатах исследований [5], опубликованных в журнале Journal of Computer and Communications, POS tagger NLTK имеет точность около 90%, что не является лучшим показателем среди теггеров. Вместо этого,