**Отчет по работе над проектом**

**“Ресурс идентификации атрибуции письменных текстов”**

**Задание:** опосредованная разметка и тестирование;

**Материал:** 8 текстов за авторством Сергея Лаврова, написанных в формальном стиле;

**Ресурс:** [http://khorom-attribution.ru/#/](http://khorom-attribution.ru/" \l "/) ;

**Атрибуты:** синтаксические сращения, конструкции с сопоставительными союзами, глагольные односоставные предложения.

Ссылка на рабочий документ: [ссылка](https://docs.google.com/document/d/1PXF29NUGPBQ1niRgxr0krrUmZKlJZ57ddL67weNCmgw/edit?usp=sharing).

Синтаксические сращения

Синтаксические сращения: формальные повторы (иду да иду), глаголы одинаковой формы (взял да и уехал).

В сравнении двух документов с Лавровым не было обнаружено ни одного случая синтаксического сращения. Чтобы выяснить, правильно ли работает алгоритм, ищем случаи СС в текстах “руками”.

В результате поисков, ни одного примера синт. сращения не было обнаружено. Исходя из этого, мы не можем говорить ни о корректности работы алгоритма, ни о неверной его работе.

Правило

1. формальные повторы: VERB + *да и/да* + VERB

2. структурные повторы:

*что было, то было*

*что было, то и есть*

*что было, то есть*

*что есть, то есть*

*что есть, то и есть*

*что есть, то и будет*

*что есть, то будет*

Ошибки

Ошибок, как и успешных результатов, не обнаружено.

Конструкции с сопоставительными союзами

Правило

1.  Сопоставительные союзы:

1. как <...> так и;
2. не только <...>, но и;
3. не столько <...>, сколько;
4. если (не) <...>, то;
5. не то что <...>, но / а;
6. не то чтобы <...> но / а;
7. в то время как;
8. между тем как;
9. по мере того как;
10. тогда как;
11. чем <...>, тем.

Ошибки

Была обнаружена всего одна конструкция с сопоставительными союзами “*Если нет, то тут ничего не поделаешь*”. Распознана она правильно.

Предложения

1. Максимальным образом расширить словарь сопоставительных союзов, включив в них и сопоставительно-противительные.

Глагольные односоставные предложения

Правило

1.Определённо-личные предложения

Глагол-сказуемое в форме 1-го или 2-го лица (нет форм прошедшего времени или условного наклонения, так как в этих формах у глагола нет лица).

pred = verb(p [“1” or “2”], incl [not “cond”], temp [not “past”])

2. Неопределённо-личные предложения

Глагол-сказуемое в форме множественного числа третьего лица (в прошедшем времени и условном наклонении глагол-сказуемое во множественном числе).

pred = verb(p [“3”], numb [“pl”])

3. Обобщённо-личные предложения

Действие конкретного лица (говорящего) является привычным, повторяющимся или представлено в виде обобщённого суждения (глагол-сказуемое стоит в форме 2-го лица единственного числа, хотя речь идёт о говорящем, то есть – о 1-м лице).

pred = verb(p [“2”], numb [“sg”])

4. Безличное предложение

1) Глагол-сказуемое в безличной форме (совпадает с формой единственного числа, третьего лица или среднего рода).

pred = verb(p [“3”], numb [“sg”], gender [“n”])

2) Составное именное сказуемое с именной частью – наречием.

pred = verb(p [“3”], numb [“sg”]) + adv

3) Сказуемое нет или глагол в безличной форме с отрицательной частицей не + дополнение в родительном падеже (отрицательные безличные предложения).

pred = {verb = “нет” or prep = “не” + verb(p [“3”], numb [“pl”]) + obj(decl [“gen”])}

5. Инфинитивные предложения

Сказуемое – независимый инфинитив.

pred = inf

Ошибки

Здесь всё более реалистично: в текстах 1-4 13 единиц, а в текстах 5-8 18 единиц.

Вопрос в том, учитываем ли мы эллиптические двусоставные предложения или нет, ведь формальное условие выполнено – в предложении из основы только глагол.

Как мы можем наблюдать, не все из извлечённых предложений отвечают критериям поиска. 3 из 31 оказались неверными. Рассмотрим каждый неправильный пример отдельно и раскроем суть ошибок алгоритма.

|  |  |  |
| --- | --- | --- |
| Тут определённо двусоставное предложение: *диалог* (подл.) *вносит вклад* (сост. сказ.) | Можем предположить, что корень ошибки в данном случае кроется в нетипичном порядке слов: именная часть сказуемого – глагольная часть сказуемого – подлежащее. | *Неоценимый вклад в общие усилия вносит регулярный, доверительный диалог на высшем уровне.* |
| СПП из двух двусоставных предложений. | Здесь вполне очевидная двусоставная основа предложения *сообщество обязано создать*. Возможно, что ошибка мотивирована формой сказуемого – СГС с основной частью среднего рода, третьего лица и единственного числа, что совпадает с формой сказуемого в безличном предложении. | *Международное сообщество обязано создать благоприятные условия для того, чтобы сами народы стран региона решали проблемы через инклюзивный национальный диалог без вмешательства извне.* |
| СПП из двух двусоставных предложений. | Опять же, причина ошибки возможно кроется в форме сказуемого придаточного предложения – средний род, третье лицо, возвратность. | *Еще один пример про те же «грабли» – вторжение в Ирак, которое в итоге завершилось появлением «Исламского государства».* |

Предложения

1. Для безличных предложений ввести список сказуемых, однозначно указывающих на односоставность (то есть бессубъектность) глагольного предложения: 1) состояние природы: *светает, холодает, смеркается*; 2) физическое или нравственное состояние человека: *знобит, тошнит, думается,* *дышится, хочется*; 3) оценка меры или необходимости какого-либо действия, источник которого остается неясным: *хватит, следует, придётся*;
2. Если в алгоритме каким-либо образом зафиксирован порядок следования отслеживаемых компонентов, то необходимо создать дополнительные условия для классификатора: обязательный поиск подлежащего (здесь отсутствие подлежащего будет флагом True), проверка сказуемого на соответствие правилам поиска.