**Отчет по работе над проектом**

**“Ресурс идентификации атрибуции письменных текстов”**

**Задание:** опосредованная разметка и тестирование;

**Материал:** 38 (19 документов по авторам-мужчинам и 19 текстов по авторам-женщинам) из публикаций в сервисах Instagram и Telegram;

**Ресурс:** <http://khorom-attribution.ru/#/> ;

**Атрибуты:** Сложные синтаксические конструкции

Ссылка на рабочий документ: [ссылка](https://docs.google.com/document/d/17leqKlAN3c0omnXkUWUYS35XobjUEuNujVMcHjKfKJY/edit?usp=sharing).

# Сложные синтаксические конструкции

## Правило

В группу (соч) входят:

соединительные союзы: и, да, а также; как… так и, мало того что... еще и, не… а, не… но, не сказать чтобы... но, не столько… сколько, не только… но и, не то чтобы… но, скорее… чем; и... и ... и; да... да... да; ни... ни... ни; ли... ли... ли; или... или ... или; то... то... то; то ли… то ли… то ли, не то... не то... не то; либо... либо... либо; будь... будь, хоть... хоть; то... то... а то; то... то... а то и; либо... либо... то ли; либо... либо... или; то ли... то ли... или; будь то... или; или... или... а может быть; может... может... а может быть; возможно... возможно; возможно... а может быть;

противительные союзы: но, да в значении но, однако, а, зато, и то;

разделительные союзы: или, либо, а то, не то, а не то; или… или, либо... либо; ли... ли, ли... или, хоть... хоть, что... что, будь то... или; а то и, а может (быть) и; не... так, если (и) не... то; а может (быть), может (быть)... может (быть), может (быть)... а может (быть); не то... не то, то ли... то ли; то... то.

В группу (подч) входят:

(1) причинные союзы (так как, потому что, поскольку, оттого что, ввиду того что, благодаря тому что, вследствие того что, в связи с тем что, в силу того что, ибо, затем что);

(2) союзы следствия (так что, а то, а не то);

(3) целевые союзы (чтобы, чтоб, для того чтобы, с тем чтобы, затем чтобы, дабы);

(4) условные союзы (если, если бы, если б, раз, ли, коль скоро, ежели (бы), коли, кабы);

(5) уступительные союзы (хотя, хоть; даром что; только бы, лишь бы; несмотря на то что, невзирая на то что; хотя бы, хоть бы, пусть, пускай; в то время как, между тем как, тогда как; добро бы, пускай бы; только, правда);

(6) временные союзы (едва, едва только, как только, как, когда, лишь, лишь только, по мере того как, после того как, с тех пор как, пока, пока не, покамест, покамест не, покуда, покуда не, прежде нежели, прежде чем, только, только что, чуть лишь, чуть, чуть только, до того как, в то время как);

(7) сравнительные союзы (как, что, будто, будто бы, как будто, как будто бы, словно (как), подобно тому как, точно, ровно (как), чем, нежели);

(8) изъяснительные союзы (что, чтобы, будто бы, как).

В группу (бесс) входят: знаки препинания “,”. “:”, “--”.

Правила:

1. Пр. содержит союзы из (соч) и (подч);
2. Пр. содержит союз или союзы из (соч) и знак препинания из (бесс);
3. Пр. содержит союз или союзы из (подч) и знак препинания из (бесс);
4. Пр. содержит союзы из (соч) и (подч) и знак препинания из (бесс).

## Виды ошибок

|  |  |
| --- | --- |
| Ошибка | Пример |
| Возвращение СПП с двумя основами | Не обольщайтесь тем, что находите. |
| Возвращение ССП с двумя основами | И кажется уже прошёл целый век, а что-то во мне продолжает жить и прорастать дарами словесности. |
| Возвращение предложений с однородными рядами | В хронологическом порядке: Годфруа Ломброзо Шопенгауэр <...> Жижек VS Лакан Прелесть свободного человека в том, что он недостижим. |
| Возвращение простых предложений со ср. оборотами | Самая большая и невежественная ошибка воспринимать "пенис" в контексте фрейдизма буквально, как орган. |

Помимо перечисленных ошибок повсеместно встречается выделение находящихся рядом слов вместе с союзами и союзными словами.

## Предложения по изменениям

1. Ввести счётчик, по которому можно будет отследить количество единиц в синтаксической конструкции. В данном случае предлагается ввести правило, отслеживающее количество “перемычек” между предложениями. Таким образом, можно будет отсечь от ССК ССП и СПП, а также бессоюзные предложения. Процесс работы представляется следующим: алгоритм ищет союзы, союзные слова и знаки пунктуации, служащие средствами бессоюзной связи в письменной речи, с помощью сравнения слов в предложении с заранее составленными списками маркеров подчинительной, сочинительной и бессоюзной связи;
2. Подсчёт основ предложений; Параллельно с подсчётом соединительных средств между основами, мы можем ввести дополнительный критерий для фильтрации предложений. Предполагается, что алгоритм должен подсчитывать количество основ в рамках одной синтаксической единицы, ограниченной рамками ССК. Так как известно, что основ должно быть более двух, то условие будет заключаться в том, что для классификации единицы как ССК необходимо наличие в ней более чем двух основ.