**Отчёт по работе над проектом**

**“Автоматическая текстовая атрибуция”**

Решение задач лингвистической разметки корпуса

Задачи работы: протестировать работоспособность функций ресурса <http://khorom-attribution.ru/#/>

* **предложения с однородными рядами (1)**;
* **предложения с обособленными приложениями (2)**;
* **целевые и выделительные обороты (3)**

на материале 200 текстов (100 текстов авторов-женщин и 100 текстов авторов-мужчин) по 500 слов жанра фанфикшен, выявить ошибки и классифицировать их, предложить путь решения неточностей.

**1) Предложения с однородными рядами**

Правила:

|  |  |
| --- | --- |
| <NOUN, NOUN, NOUN>  <ADJF, ADJF, ADJF>  <ADJS, ADJS, ADJS>  <VERB, VERB, VERB>  <INFN, INFN, INFN>  <и NOUN, и NOUN, и NOUN>  <NOUN, NOUN и NOUN>  <NOUN, и NOUN, и NOUN>  <NOUN, NOUN да NOUN>  <NOUN, NOUN да и NOUN>  <да и NOUN, да и NOUN, да и NOUN>  <NOUN, да и NOUN, да и NOUN>  <NOUN, да NOUN, да NOUN>  <ни NOUN, ни NOUN, ни NOUN>  <NOUN, также NOUN, также NOUN>  <NOUN, тоже NOUN, тоже NOUN>  <NOUN, также и NOUN, также и NOUN>  <NOUN, а NOUN, NOUN>  <NOUN, NOUN, а NOUN>  <NOUN и NOUN, а NOUN>  <NOUN, но NOUN, NOUN>  <NOUN, NOUN, но NOUN>  <NOUN и NOUN, но NOUN>  <NOUN, или NOUN, или NOUN>  <NOUN, NOUN или NOUN>  <или NOUN, или NOUN, или NOUN>  <NOUN, либо NOUN, либо NOUN>  <NOUN, NOUN либо NOUN>  <либо NOUN, либо NOUN, либо NOUN>  То же для ADJF, ADJS, VERB, INFN, PRTF, PRTS, GRND, PRED | <PRTF, PRTF>  <PRTS, PRTS>  <GRND, GRND>  <PRED, PRED>  <и PRTF, и PRTF>  <PRTF и PRTF>  < PRTF да PRTF>  <PRTF да и PRTF>  <да и PRTF, да и PRTF>  <ни PRTF, ни PRTF>  <PRTF, также PRTF>  <PRTF, тоже PRTF>  <также и PRTF, также и PRTF>  <PRTF, также и PRTF>  <PRTF, а PRTF>  <PRTF, но PRTF>  <или PRTF, или PRTF>  <PRTF или PRTF>  <либо PRTF, либо PRTF>  <PRTF либо PRTF>  То же для NOUN, ADJF, ADJS, VERB, INFN |

|  |  |  |
| --- | --- | --- |
| **№** | **Пример ошибки** | **Объяснение** |
| **1** | 1. *Ладно женское внимание, он сам не без греха, Ксиу и Ланиинг, дочери Линтао, были далеко не только служанками, но и наложницами, но десяток?*  2. *В глаза сразу ударил свет магических фонарей, стены и потолок были затянуты лиловой и нежно-розовой тканью, а прямо посредине неплохого такого зала стояло несколько мягких диванчиков и стол с разными яствами: печеньем, тортиками и всевозможными сладостями, манящими взгляд... Хе-хе, здесь даже камин был!*  3. *Это был каменный лабиринт, стены и пол которого покрывали растения, скрывающие все ловушки и опасности этого страшного места.*  4. *Мать при этом застыла в шоке и ужасе, а Гарри зачарованно рассматривал маленькое, медленно шевелящее крылышками огненное чудо.*  5. *Не монстра, блуждающего по тёмному лабиринту в поисках знаний и силы, а отца.*  6. *Аколиты встали на дальних концах, Роши и Тарджи — на боковых.* | Выделение трёх членов предложения, из которых только два составляют однородный ряд:   * один из членов предложения выражен именем нарицательным, обозначающим неодушевлённый объект, остальные два — именами собственными ЛИБО один из них выражен именем собственным, а другие два — именами нарицательными, обозначающими неодушевлённые объекты; * члены предложения выражены нарицательными именами существительными, которые являются неодушевлёнными и одушевлёнными; * два члена предложения выражены существительными в одном падеже, а третий — существительным в другом падеже; * члены предложения, которые могут быть выражены существительными в одном и том же падеже, входят в разные части ССП / СПП. |
| **2** | 1. *В нашем с Котори доме довольно просторная кухня, всё современное – созданное для облегчения жизни: большая белая микроволновка, посудомойка, электроплита, электрическая вытяжка, огромный холодильник, явно не рассчитанный на двух человек, проживающих в довольно большом доме.*  2. *- Омлет с яйцами, молоком, мясом и... креветками, м?*  3. *В обширных знаниях Тома существовал как минимум один вариант объяснения этому: у кого-то, скорее всего у старого долькомана, имеется специальный артефакт наподобие часов в кухне Уизли, настроенный конкретно на Гарри Поттера, который показывает его состояние: жив, здоров, болен, в опасности, мертв.*  4. *Джеймс привык к тому, что Рождество — это по-настоящему волшебный, уютный, тёплый и семейный праздник.*  5. *– Трусы, косметика, – отмахнулся преподаватель, – спички, керосин, томик Виллиана Экспера, откуда я знаю, что вы причисляете к предметам первой необходимости?*  6. *Тогда умерли бы и Фрэнк, и Алиса, и Невилл и даже Августа.* | Выделение только части однородного ряда, что происходит вследствие того, что система ищет однородные ряды, состоящие или из двух, или из трёх членов предложения. |
| **3** | 1. *Катарина бесстыжая, абсолютно бессовестная и острая на язык, Катарина-держи-друзей-близко-а-врагов-в-неведении, Катарина всегда выглядит так, как будто видит тебя насквозь, и если она осмеливается вот так вот дерзить преподавателю, стоит ли уточнять, как она разговаривает со сверстниками?*  2. *Для начала он выставил на обеденный стол стакан с водой, пёрышко, подсвечник со свечой и небольшой горшок с землёй, в который зарыл боб.*  3. *– возразила я, зачем-то вцепившись в его плечи – то ли с подоконника упасть боялась, то ли думала, может, убежит куда – всё-таки, у сестрёнки размер бюста аж третий с половиной, а мне до неё, увы, размера не хватает.*  4. *– И это должен быть рюкзак, Лина, рюкзак, а не чемодан, – он внимательно посмотрел на весёлую рыжеволосую Дженнон и добавил: – И он должен быть меньше тебя.*  5. *– Убери листик, Май, перекличку проведём устно.*  6. *— Дамы и господа, жители и гости нашего прекрасного города!* | Выделение членов предложения, не составляющих однородный ряд: система ищет три любых перечисляемых через запятую члена предложения, имеющих одинаковую частеречную принадлежность, в том числе “захватывает” вводные слова (*может*, омонимичное глаголу) и обращения, не являющиеся членами предложения. |
| **4** | 1. *Именно для этого существуют маг-кластеры, самые крупные из которых — Европейский, Американский, Латинский, Русско-Азиатский, Китайский и Японский, которые соответствуют народам, жившим там.*  *Именно для этого существуют маг-кластеры, самые крупные из которых — Европейский, Американский, ЛатинскийАмериканский, Латинский, Русско-Азиатский, Китайский и Японский, которые соответствуют народам, жившим там.*  *Именно для этого существуют маг-кластеры, самые крупные из которых — Европейский, Американский, ЛатинскийАмериканский, Латинский, Русско-АзиатскийЛатинский, Русско-Азиатский, Китайский и Японский, которые соответствуют народам, жившим там.*  *2. У магглов там регистрируют факты рождения, смерти, свадьбы, усыновления и опекунства?*  *У магглов там регистрируют факты рождения, смерти, свадьбысмерти, свадьбы, усыновления и опекунства?*  *У магглов там регистрируют факты рождения, смерти, свадьбысмерти, свадьбы, усыновлениясвадьбы, усыновления и опекунства?*  3. *Много достопримечательностей, архитектура, музеи, театры... банки.*  *Много достопримечательностей, архитектура, музеиархитектура, музеи, театры... банки.*  4. *После смерти бабушки Джеймс думал, что Рождество никогда не станет прежним, но мама словно переняла эстафету и старалась, чтобы эти зимние праздники были хорошими, они тоже были наполнены теплом, весельем, гостями, горячим какао у камина и захватывающими историями.*  *После смерти бабушки Джеймс думал, что Рождество никогда не станет прежним, но мама словно переняла эстафету и старалась, чтобы эти зимние праздники были хорошими, они тоже были наполнены теплом, весельем, гостямивесельем, гостями, горячим какао у камина и захватывающими историями.*  5. *Конкурс проходил в несколько этапов: для художников, портных, кузнецов, дизайнеров.*  *Конкурс проходил в несколько этапов: для художников, портных, кузнецовпортных, кузнецов, дизайнеров.* | Техническая ошибка программы: одно и то же предложение выводится на экран несколько раз, в каждом последующем из которых дублируется часть предложения и “захватывается” всё больше членов предложения, в результате чего программа неправильно считает количество найденных однородных рядов. |
| **5** | 1. *— В девять вечера мистер Бэгмэн, судья и устроитель Турнира, будет ждать участников Турнира на школьном поле для квиддича.*  2. *Вообще она девочка, Муфлоша, Муфточка, Муфлень – жница извращается над именем упрямой камелиды как может.* | Выделение трёх членов предложения, один из которых является определяемым словом, а остальные — приложениями к нему. |

**Предложения по исправлению и улучшению**:

1. При выделении однородных подлежащих и дополнений, выраженных существительными, учитывать:

* категорию одушевлённости и их отнесённость к именам собственным и нарицательным;
* падеж;
* часть ССП / СПП, к которой они относятся.

1. При выделении однородных сказуемых, выраженных глаголами, учитывать их морфологическое время.
2. Не выделять обращения, вводные слова, омонимичные соответствующим глаголам (*может*, *казалось* и др.) и другие конструкции, которые членами предложения не являются.
3. Искать однородные ряды, длина которых больше трёх. При этом необходимо также помнить, что однородные члены могут быть:

* распространены согласованными (*тёплый и семейный праздник*) и несогласованными определениями (*стакан с водой*);
* иногда по-разному выражены морфологически (*показывает его состояние: жив, здоров, болен, в опасности, мертв*);
* разделены словами автора (*– Трусы, косметика, – отмахнулся преподаватель, – спички, керосин, томик Виллиана Экспера, откуда я знаю, что вы причисляете к предметам первой необходимости?*).

Что касается последнего однородного члена предложения, он может присоединяться союзом *и* с усилительной частицей *даже* (*Тогда умерли бы и Фрэнк, и Алиса, и Невилл и даже Августа*); отделяться многоточием (*- Омлет с яйцами, молоком, мясом и... креветками, м?*).

1. Однородными членами не являются:

* повторение какого-либо члена предложения: *Костры, костры, костры, вся степь в кострах (В. Тендряков)*; *Босая, на жнивье колючем, все плачет, плачет мать (К. Симонов)*;
* фразеологизированные сочинительные сочетания: *и так и сяк*, *и смех и грех*, *ни с того ни с сего*, *ни за что ни про что*, *ни рыба ни мясо*, *ни сном ни духом*, *рано или поздно*, *ни жив ни мертв*;
* простые осложнённые сказуемые (сочетание двух глаголов, первый из которых имеет значение волеизъявления, побуждения к действию, названному вторым глаголом): *возьму и скажу*, *взял да и пожаловался*, *пошёл и отказался*;
* уточняющие члены предложения: *Однажды, осенью, возвращаясь с отъезжего поля, я простудился (И. Тургенев)*.

1. Добавить в правила однородные наречия (ADVB) и компаративы (COMP).
2. Искать однородные, соединяющиеся и другими союзами: соединительными *ни – ни*, разделительными *то – то, не то – не то*, противительным *да*, двойными *как…, так и*, *не только…, но и*, *хотя и…, но*.

**2) Предложения с обособленными приложениями**

Правила:

Формула (приложение отделено двумя тире):

1) <…> **NOUN** (в nomn, gent, datv, accs, ablt, loct, voct, gen1, gen2, acc2, loc1, loc2 в sing или plur) – **NOUN** в том же падеже, что и первое – <…>

2) <…> **NOUN** (в nomn, gent, datv, accs, ablt, loct, voct, gen1, gen2, acc2, loc1, loc2 в sing или plur) – **NOUN** в том же падеже, что и первое + POST[[1]](#footnote-1) – <…>

3) <…> **NOUN** (в nomn, gent, datv, accs, ablt, loct, voct, gen1, gen2, acc2, loc1, loc2 в sing или plur) – POST[[2]](#footnote-2) + **NOUN** в том же падеже, что и первое – <…>

5) <…> **NOUN** (в nomn, gent, datv, accs, ablt, loct, voct, gen1, gen2, acc2, loc1, loc2 в sing или plur) – то есть / или / как + **NOUN** в том же падеже, что и первое – <…>

6) <…> **NOUN** (в nomn, gent, datv, accs, ablt, loct, voct, gen1, gen2, acc2, loc1, loc2 в sing или plur) – то есть / или / как + **NOUN** в том же падеже, что и первое + POST

7) <…> **NOUN** (в nomn, gent, datv, accs, ablt, loct, voct, gen1, gen2, acc2, loc1, loc2 в sing или plur) – то есть / или / как + POST + **NOUN** в том же падеже, что и первое – <…>

8) <…> **NOUN** (в nomn, gent, datv, accs, ablt, loct, voct, gen1, gen2, acc2, loc1, loc2 в sing или plur), то есть / или / как + **NOUN** в том же падеже, что и первое, <…>

9) <…> **NOUN** (в nomn, gent, datv, accs, ablt, loct, voct, gen1, gen2, acc2, loc1, loc2 в sing или plur), то есть / или / как + **NOUN** в том же падеже, что и первое + POST, <…>

10) <…> **NOUN** (в nomn, gent, datv, accs, ablt, loct, voct, gen1, gen2, acc2, loc1, loc2 в sing или plur), то есть / или / как + POST + **NOUN** в том же падеже, что и первое, <…>

|  |  |  |
| --- | --- | --- |
| **№** | **Пример ошибки** | **Объяснение** |
| **1** | а) 1. *Меня больше интересовал вопрос, как Стражи, что вроде как их «чуют», умудрились пролюбить и диверсантов, и подкоп.*  2. *Я до того места, как она в мульте появилась, не досмотрел, но мне Серёга рассказывал и показывал «куски классного мочилова», чтобы заинтересовать.*  б) 1. *Я была совершенно ошалевшая от внезапно привалившего счастья, как кошка в бидоне валерьянки, которая тонет, охренев от радости.*  2. *А теперь Джирайя, как гончий пёс, искал малейшие следы бывшего сокомандника и не мог отделаться от чувства, что подбирается к нему всё ближе.*  3. *Скорее всего просто путник, как и я. Настораживает лишь отсутствие лошадей.*  в) *Дойдя до кафедры, она повернулась в сторону класса, повернулась неожиданно массивно и всем телом, как если бы была вросшим в пол аудитории деревом, которое зачем-то научилось передвигаться.*  г) *Но работа, как известно, не волк, работа хуже, поэтому Цефея натянула на себя вежливую улыбку и отправилась в указанную сторону.*  д) *Стоило шагнуть ближе к кровати, как он открыл глаза — внимательные, без следа сонной мути.*  е) *А в тот день только услышала из коридора, как наши «батонки» обсуждали, что приз в тысячу галлеонов вручили Гарри Поттеру, а завтра утром, ещё до завтрака, карета улетит во Францию.*  ж) *Если бы я была художницей, мне бы в голову пришли сотни сравнений – цвет морской волны или неправильного изумруда, отливающего синим, или цвет бедра испуганной нимфы – один чёрт знает, что это вообще за цвет.*  з) *Хм, ну и внешняя подпитка, как магией, так и просто энергией - в ходе преобразований, особенно ученических, я улавливал изменения температуры в комнате, точнее, понижение где-то на градус-полтора - для обычного человека совсем незаметно.*  и) *Может быть, но тебе не кажется странным, что мы, по сути, как раз сражаемся с теми, кого Создатель, опять же, согласно мнению церкви, наслал на этот мир?*  к) *— в лоб спросила Сенджу, как только техника замкнулась.*  л) *- Что за вирус, каким образом он проник на базу, как именно воздействует на систему?*  м) *Перемены характера Чана-младшего на его пятнадцатилетие были просто подарком Духов, но вот то, к чему эти перемены привели… С одной стороны, как солдат, он гордился славным представителем своего рода, но вот как отец, раз за разом получая доклады о новых столкновениях Вестника с превосходящими силами противника, испытывал страх за сына.* | Программа “захватывает” любые конструкции, включающие **NOUN, *как***, и выделяет предложение до конца:  а) определяемое NOUN в главном предложении + придаточное определительное с *как*;  б) NOUN + сравнительный оборот с *как*;  в) NOUN + придаточное сравнительное с *как*;  г) NOUN + вводная конструкция с *как*;  д) NOUN + *как* в составе фразеологизированной конструкции *стоило..., как* со значением обусловленности;  е) NOUN + придаточное изъяснительное с *как*;  ж) обобщающее слово NOUN + ряд однородных членов;  з) NOUN + *как* в составе союза *как…, так и*;  и) NOUN + *как* в составе частицы *как раз*;    к) NOUN + *как* в составе союза *как только* в придаточном времени;  л) NOUN + вопросительное местоименное наречие *как*;  м) NOUN + приложение с *как*. |
| **2** | 1. *Выражение на лице почти каменно-спокойное, будто он вообще изваяние – хочется потыкать пальцем в щёку, чтоб убедиться, что живой – но это у меня воображение живое слишком, другие, наверное, привыкли.*  2. *Или кто-то из родичей скоропостижно помер, отписав ему... нет, не наследство — дождешься от демонов, аж десять раз — парочку проклятий помощнее.* | Выделение NOUN + вставная конструкция. |
| **3** | 1. *И я, на самом деле, даже и не знала, куда пялиться – на душку Шиффера, или на этого чудика в одном носке и всё тех же спортивных штанах, который не мог найти даже джинсы.*  2. *И оправдано, и звучит ничего так. Как в старых детективах, или в ковбойских сериалах.* | Неверная авторская пунктуация: перед единичным союзом *или* запятая не нужна, так как он соединяет однородные члены предложения. |
| **4** | 1. *Умница Габриель догадалась принести мне в Больничное крыло, из которого с утра пораньше меня выписала медичка, одежду, то есть «мантию» или «школьную форму».*  2. *Требовалось составить списки, утвердить надписи, подготовить что-то вроде магических эскизов, то есть волшебных иллюзий, и отправить всё это изготовителям-кузнецам.* | Выделение NOUN + пояснительный член предложения с союзом *то есть*. |
| **5** | а) *Для такого уровня тактичности надо быть или конченым мудаком, или вкрай отбитым пофигистом.*  б) *— детям запрещено давать это зелье, или вам это неизвестно, мистер Поттер?* | Программа “захватывает” любые конструкции, включающие **NOUN, *или***, и выделяет предложение до конца:  а) NOUN + повторяющийся союз *или*, соединяющий однородные члены предложения;  б) NOUN + одиночный союз *или*, соединяющий части ССП. |

**Предложения по исправлению и улучшению:**

1. Исключить из правил звательный падеж (voct).
2. Искать также не присоединяемые союзами приложения (одиночные / распространённые), обособленные запятыми.
3. Обособленные приложения с *как* выделять, если они стоят до определяемого слова (дополнительное значение причины / уступки): *Как человек образованный, наш дедушка хорошо знал историю*. В этом случае возможна омонимия с другими конструкциями с *как*, но ошибочных примеров должно быть меньше.
4. Исключить конструкции с союзом *то есть*, являющимся маркером пояснительного члена предложения.
5. Любое приложение должно состоять из существительного в том же падеже и числе, что и определяемое слово, и, если приложение распространённое, также из согласованного или несогласованного определения.
6. Выделять приложения со словами *по имени*, *по фамилии*, *по прозвищу*, *по происхождению*, *родом* и др., если они обособлены.
7. В случае с именами собственными выделять обособленное приложение:

* если оно стоит до или после имени собственного: *Упрямец во всём, Илья Матвеевич оставался упрямцем и в учении (Кочетов)*; *И. Левитан, известный художник, тонко передаёт игру светотени в берёзовой роще*;
* если оно выражено именем собственным и стоит после имени нарицательного: *В разговор изредка вставляет слово Любина тётка, Ксения Фроловна Горина (Песков)*.

1. В случае с именами нарицательными выделять обособленное приложение:

* если оно распространено и стоит до и после определяемого слова: *Старуха, Гришкина мать, умерла, но старики, отец и тесть, были ещё живы (Салтыков-Щедрин)*; *Несчастью верная сестра, надежда в мрачном подземелье разбудит бодрость и веселье (Пушкин)*;
* если оно одиночное и стоит после определяемого слова: *Ухаживала за мной одна девушка, полька (Горький)*;
* если оно единичное и определяемое слово имеет при себе пояснительные слова: *Тут на широкой улице встретился им повар генерала Жукова, старичок (Чехов)*.

1. Выделять приложения, которые относятся к личному местоимению, так как они обособляются независимо от его местонахождения в предложении. Приложение должно быть выражено существительным в том же падеже и числе, что и местоимение, с согласованным / несогласованным определением, если приложение распространённое.
2. Выделять одиночное или распространённое согласованным / несогласованным определением приложение, обособленное тире и стоящее в конце и середине предложения: *Давно замечено, что одиноко растёт, не подпуская никого селиться возле себя, одно только дерево – берёза*; *В комнату вбежала оживлённая малышка – любимица отца – в розовом платье, с огромным бантом на макушке.* Приложение должно быть выражено существительным в том же падеже и числе, что и определяемое слово, и не должно содержать личные формы глагола, частицы и пр., что может входить во вставные конструкции.
3. При поиске приложений с союзом *или*:

* перед союзом должна стоять запятая;
* союз должен быть одиночным;
* союз должен присоединять существительное в том же числе и падеже, что и определяемое слово, стоящее перед *или*, с согласованным / несогласованным определением, если приложение распространённое.

1. Учитывать границы выделяемого приложения, обращать внимание на пунктуацию.

**3) Целевые и выделительные обороты**

Правила:

1) Cтруктуры типа: с целью / из расчёта + INFN

2) Предлоги: кроме, помимо, включая, наряду с

|  |  |  |
| --- | --- | --- |
| **Пример ошибки** | **Объяснение** | **Предложение по исправлению** |
| 1. *Кроме того, в соответствии с направлением передачи силы, аколиты положили правые руки на спины сатьянам, а те — на лопатки мне.*  2. *Через четыре месяца… — я уже узнала дату, и неприятным для меня известием, что, кроме всего прочего, я ещё и переместилась во времени, так как в данном сказочном мире шёл не август две тысячи восемнадцатого, а февраль тысяча девятьсот девяносто пятого.*  3. *Гарен в академии Дюрандаль сам учился, знает, какая эта академия конченая по всем фронтам; что вообще ещё можно сказать об учебном заведении, в котором, помимо прочего, выращивают убийц?*  4. *А, кроме того, я абсолютно несексуальная девственница, которая не готова к серьёзным в интимном плане отношениям.* | Выделение предлогов *кроме*, *помимо*, которые являются частью не выделительного оборота, а вводной конструкции | Не учитывать при нахождении выделительных оборотов конструкции *кроме того*, *но кроме того*, *и кроме того*, *кроме этого*, *кроме всего прочего*, *кроме шуток*, *помимо того*, *помимо этого*, *помимо всего*, *помимо всего прочего*, которые относятся к вводным.  Но! Необходимо выделять союз *кроме того, что*, так как он вводит выделительный оборот, как в примере:  *Ничего примечательного в нём не содержалось — ну, кроме того, что Джи как-то по пьяни поэкспериментировал на нём в фуин, благодаря чему верной тряпке сносу не было.* |

1. Скорее всего, здесь будет либо ADJF в том же падеже, что и NOUN, либо PRTF в том же падеже, что и NOUN [↑](#footnote-ref-1)
2. Скорее всего, здесь будет либо ADJF в том же падеже, что и NOUN, либо PRTF в том же падеже, что и NOUN [↑](#footnote-ref-2)