

После сравнения ряда примеров(см. Приложение 1) из корпуса на разные конструкции, мы получили такой результат:

Таблица различий работы парсеров.

Таблица 1.

| Различие | UDPipe | CoreNLP | Spacy |
|--|---|---|---|
| 1. Случаи, где предикат не является глаголом | предикат - root вспомогательный глагол - cop | предикат - root вспомогательный глагол - cop | предикат - attr вспомогательный глагол - root |
| 2. Отсутствие копулы | справляется, правильно отмечает подлежащее и сказуемое | справляется, правильно отмечает подлежащее и сказуемое | не справляется, нужно проверить |
| 3. POS копулы (is/are) | AUX | VERB | AUX |
| 4. однородные члены + копула, которая не соответствует числу (5.4) | + | + | - |
| 5. вводные конструкции в начале (to sum up) | ставит конструкцию в root | в advcl | в advcl |
| 6. однородные подлежащие через or с глаголом в единственном числе | не определяет подлежащие как однородные | все ок | все ок |
| 7. neither...nor с ошибкой | ? | все ок | neither относится к подлежащему |
| 8. he or she | проблемы | все ок | все ок |
| 9. знак процента как подлежащее | все рушится | все ок | все ок |
| 10. национальности как подлежащее (the French) | не понимает, что French может быть подлежащим | все ок | все ок |
| 11. герундий как подлежащее | вроде все ок? | неправильно определяет если неправильное согласование | неправильно определяет если неправильное согласование |
| 12. "the differents" | - | - | + |
| 13. к чему относится relative clause (пункт 11) | к ближайшему существительному? | к ближайшему? | к первому? |

Table 1.

| Различие | UDPipe | CoreNLP | Spacy |
|----------------------------|-----------------|------------------|-----------------|
| 1. Случаи, где предикат не | предикат - root | predicate - root | предикат - attr |

| | | | |
|--|--|---|---|
| является глаголом | вспомогательный глагол - cop | copula - cop | вспомогательный глагол - root |
| 2. Отсутствие копулы | справляется, правильно отмечает подлежащее и сказуемое | Works well, correctly marks the subject and the predicate | не справляется, нужно проверить |
| 3. POS копулы (is/are) | AUX | VERB | AUX |
| 4. однородные члены + копула, которая не соответствует числу (5.4) | + | + | - |
| 5. вводные конструкции в начале (to sum up) | ставит конструкцию в root | Marks introductory words as advcl | в advcl |
| 6. однородные подлежащие через or с глаголом в единственном числе | не определяет подлежащие как однородные | Works well | все ок |
| 7. neither...nor с ошибкой | ? | Works well | neither относится к подлежащему |
| 8. he or she | проблемы | Works well | все ок |
| 9. знак процента как подлежащее | все рушится | Works well | все ок |
| 10. национальности как подлежащее (the French) | не понимает, что French может быть подлежащим | Works well | все ок |
| 11. герундий как подлежащее | вроде все ок? | Can sometimes incorrectly mark the gerund subject if the agreement is wrong | неправильно определяет если неправильное согласование |
| 12. "the different" | - | - | + |
| 13. к чему относится relative clause (пункт 11) | к ближайшему существительному? | Relative clause is connected to the nearest noun | к первому? |

Далее мы собрали несколько новых необходимых примеров (см. Приложение 2) случаев, в которых у нас возникли вопросы.

Таблица 2.

| Sentence | UDPipe | Stanford | Spacy |
|------------------|--|---|--|
| NO COPULA | beautiful(<i>advcl</i>), she(<i>nsubj</i>) - <i>прилагательное</i> | she(<i>nsubj</i>) really beautiful(<i>advcl</i>) все ок, <i>прилагательное</i> | beautiful(<i>advcl</i>), she(<i>nsubj</i>) - <i>прилагательное</i> |

| | | | |
|--|---|--|--|
| <p>Naomy is expected to be nominated as the winner of competition between the models, because she really beautiful. - esl_01293.txt</p> <p>Naomy is expected to be nominated as the winner of competition between the models, because she is really beautiful. - esl_01293.txt</p> | <p><i>является вершиной зависимой клаузы</i></p> <p><i>she(nsubj), is(cop), beautiful(advcl) - прилагательное - вершина зависимой клаузы, копула зависит от прилагательного</i></p> | <p><i>является вершиной зависимой клаузы</i></p> <p><i>she (nsubj) is (cop) really beautiful (advcl) все ок, прилагательное является вершиной зависимой клаузы</i></p> | <p><i>является вершиной зависимой клаузы</i></p> <p><i>she(nsubj), is(advcl), beautiful(acomp) - копула - вершина зависимой клаузы, именная часть сказуемого зависит от копулы</i></p> |
| <p>NO COPULA</p> <p>In contrast, other people claim that facilities not so important.</p> <p>In contrast, other people claim that facilities are not so important.</p> | <p><i>that (mark) facilities (nsubj) not so important (ccomp) вершина important зависит от root claim</i></p> <p><i>that (mark) facilities (nsubj) are (cop) not so important (ccomp) вершина important зависит от root claim</i></p> | <p><i>that (dobj) facilities (dep) not so important (advmod) не понял, что это зависимая клауза, поставил that в прямой объект и просто прикрепил к нему все остальное</i></p> <p><i>that (mark) facilities (nsubj) are (cop) not so important (ccomp) все ок, прилагательное является вершиной зависимой клаузы</i></p> | <p><i>important(ccomp) - вершина зависимой клаузы, facilities(nsubj)</i></p> <p><i>are(ccomp) - вершина зависимой клаузы, important(acomp), facilities(nsubj)</i></p> |
| | <p>В примерах, где копула присутствует, все три парсера разметили предложение правильно, следуя своей логике разметки.</p> <p>В примере, где копула отсутствует, у UDPipe ничего не изменилось, только пропала копула, Срасу переместил прилагательное в вершину клаузы вместо копулы, а Stanford CoreNLP не справился с примером, так как не распознал зависимую клаузу.</p> | | |
| <p>NO COPULA</p> <p>It is shown in the picture below.</p> <p>It shown in the picture below.</p> | <p><i>it (nsubjpass) is (auxpass) shown (root) все хорошо</i></p> <p><i>it (root) shown (advcl) in the picture below. не справился: разделил</i></p> | <p><i>it (nsubjpass) is (auxpass) shown (root) correct</i></p> <p><i>it (nsubj) shown (root) не понял, что пассив, в остальном ок</i></p> | <p><i>it (nsubjpass), is (auxpass), shown(root, verb)</i></p> <p><i>it(nsubj) shown (root, verb) не понял, что пассив, в остальном ок</i></p> |

| | | | |
|--|---|--|--|
| | <p>на две клаузы с нелогичным <i>it-root</i></p> | | |
| <p>PUNCTUATION INFLUENCE Componential analysis, or lexical decomposition, is method of presenting the sense of lexeme through it parts.</p> <p>Componential analysis or lexical decomposition is method of presenting the sense of lexeme through it parts.</p> | <p>ничего не меняется, все распознано верно</p> | <p>Componential analysis (<i>nsubj</i>), or (<i>cc</i>) lexical decomposition (<i>conj</i>), is (<i>cop</i>) method (<i>root</i>)</p> <p>Componential analysis (<i>nsubj</i>) or (<i>cc</i>) lexical decomposition (<i>conj</i>) is (<i>cop</i>) method (<i>root</i>) <i>одинаково разметил</i></p> | <p>Componential analysis (<i>nsubj</i>), or (<i>cc</i>) lexical decomposition (<i>conj</i>), is (<i>root-AUX</i>) method (<i>attr npu is-root</i>)</p> <p>Componential analysis (<i>nsubj</i>), or (<i>cc</i>) lexical decomposition (<i>conj</i>), is (<i>root-AUX</i>) method (<i>attr npu is-root</i>) <i>одинаково разметил</i></p> |
| <p>PUNCTUATION INFLUENCE 73% of them prefer facebook, 25% use Instagram and 31% go to LinkedIn.</p> <p>73% of them prefer facebook, 25% use Instagram, and 31% go to LinkedIn.</p> | <p>73% (<i>nsubj</i>) of them prefer (<i>root</i>) facebook, 25% (<i>conj</i>) use (<i>mark</i>) Instagram (<i>nsubj</i>) and 31% (<i>conj</i>) go (<i>advcl</i>) to LinkedIn. <i>Первая клауза выделена правильно, вторая и третья -- совершенно нет. Парсер увидел 4 клаузы, втч просто 31% и просто 25%.</i></p> <p>73% (<i>nsubj</i>) of them prefer (<i>root</i>) facebook, 25% (<i>nsubj</i>) use (<i>conj</i>) Instagram (<i>obj</i>), and 31% (<i>nsubj</i>) go (<i>conj</i>) to LinkedIn. <i>В случае с запятой все определено совершенно верно.</i></p> | <p>73% (<i>nsubj</i>) of them prefer (<i>root</i>) facebook, 25% (<i>amod к Instagram</i>) use (<i>compound к Instagram</i>) Instagram (<i>nsubj к go</i>) and 31% (<i>conj к go</i>) go (<i>parataxis, зависит от root</i>) to LinkedIn. <i>Первая клауза выделена правильно. Далее он посчитал, что 25% и use -- модификаторы к Instagram; Instagram и 31% -- подлежащие к паратаксису go</i></p> <p>73% (<i>nsubj</i>) of them prefer (<i>root</i>) facebook, 25% (<i>amod к Instagram</i>) use (<i>compound к Instagram</i>) Instagram</p> | <p>73% (<i>%-nsubj</i>) of them prefer (<i>ccomp, зависит от use</i>), 25% (<i>% - nsubj от use</i>) use (<i>root</i>) Instagram (<i>dobj от use</i>) and 31% (<i>nsubj к go</i>) go (<i>conj, зависит от root</i>) to LinkedIn.</p> <p>73% (<i>%-nsubj</i>) of them prefer (<i>root</i>), 25% (<i>% - nsubj от use</i>) use (<i>xcomp, зависит от prefer-root</i>) Instagram (<i>dobj от use</i>) and 31% (<i>nsubj к go</i>) go (<i>conj, зависит от prefer-root</i>) to LinkedIn.</p> |
| | <p>Присутствие или отсутствие запятых не влияет на разметку этого предложения тремя парсерами.</p> | | |

| | | | |
|--|--|--|--|
| | | <p><i>(dobj к prefer) and 31% (nsubj) go (conj, зависит от Instagram) to LinkedIn.</i></p> <p><i>Первая клауза выделена правильно. Решил что Instagram -- прямой объект к prefer, 25% и use -- модификаторы к Instagram. Третью клаузу обозначил правильно, но прикрепил ее не к root, а к Instagram</i></p> | |
| | <p>Предложение, в котором запятая перед 'and' отсутствует было размечено лучше всего парсером Spacy, который распознал все три клаузы. Его единственный ошибкой было то, что он посчитал первую клаузу компонентом ко второй клаузе. UDPipe и Stanford CoreNLP правильно распознали первую клаузу, но ошибочно разметили две другие.</p> <p>Предложение, где все запятые проставлены, было размечено лучше всего парсером UDPipe, который правильно разобрался со всеми клаузами и их отношениями. Spacy допустил ошибку в том, что посчитал вторую клаузу открытым компонентом к первой, но в остальном тоже разметил предложение правильно. Stanford CoreNLP не справился с задачей даже при правильной постановке запятых, поскольку посчитал некоторые элементы объектами или модификаторами, хотя они таковыми не являются.</p> | | |
| <p>PUNCTUATION INFLUENCE</p> <p>As we can see, the most popular place among the foreigners is North America, the number of visitors have increased since 1990 till 2000, and decreased a little bit to 2005.</p> <p>As we can see, the most popular place among the foreigners is North America, the number of visitors have increased since 1990 till 2000 and</p> | <p>ничего не меняется, вторая клауза разделяется и не распознается <i>nsubj</i> place</p> | <p>the most popular place (<i>nsubj</i>) among the foreigners is (<i>cop</i>) North America (<i>root</i>), the number (<i>nsubj</i>) of visitors have (<i>aux</i>) increased (<i>parataxis</i>) since 1990 till 2000, and (<i>cc</i>) decreased (<i>conj</i>) a little bit to 2005.</p> <p><i>increased -- паратаксис к root</i></p> <p><i>decreased -- однородный с increased, присоединяется к increased</i></p> | <p>the most popular place (<i>nsubj</i>) among the foreigners is (<i>ccomp om increased-root</i>) North America (<i>attr om is</i>), the number (<i>nsubj om increased-root</i>) of visitors have (<i>aux</i>) increased (<i>root</i>) since 1990 till 2000, and (<i>cc</i>) decreased (<i>conj om increased</i>) a little bit to 2005. - в обоих случаях</p> <p><i>Наличие запятой не повлияло на правильное понимание структуры всего предложения.</i></p> |

| | | | |
|---|--|---|--|
| <p>decreased a little bit to 2005.</p> | | <p>the most popular place (<i>nsubj</i>) among the foreigners is (<i>cop</i>) North America (<i>root</i>), the number (<i>nsubj</i>) of visitors have (<i>aux</i>) increased (<i>parataxis</i>) since 1990 till 2000 and (<i>cc</i>) decreased (<i>conj</i>) a little bit to 2005. <i>increased -- паратаксис к root</i> <i>decreased -- однородный с increased, присоединяется к increased</i></p> | |
| <p>Наличие или отсутствие запятой не сильно повлияло на работу парсеров. Stanford CoreNLP разметил все правильно, посчитав второй глагол паратаксисом от первого, правильно обнаружил однородные члены. Спасу решил, что первая клауза является компонентом ко второму глаголу, а UDPipe не смог обнаружить подлежащее первой клаузы.</p> | | | |
| <p>There are apples and pineapples on the table.</p> <p>There are an apple and pineapples on the table.</p> <p>There is apples and pineapples on the table.</p> <p>There is an apple and pineapples on the table.</p> | <p>Все как и ожидалось</p> | <p>Все ок. apple(s) - <i>nsubj</i> is/are - <i>root</i></p> | <p>are(<i>root</i>), apples(<i>attr</i>), pineapples(<i>conj</i>, зависит от apples)</p> <p>are(<i>root</i>), apple(<i>attr</i>), pineapples(<i>conj</i>, зависит от apple)</p> <p>is(<i>root</i>), apples(<i>attr</i>), pineapples(<i>conj</i>, зависит от apples)</p> <p>is(<i>root</i>), apple(<i>attr</i>), pineapples(<i>conj</i>, зависит от apples)</p> |
| <p>Во всех примерах парсеры справились с задачей. Отсутствие или присутствие артикля, правильное или неправильное согласование не повлияли на их работу.</p> | | | |
| <p>VERBS MISTAKEN FOR NOUNS When a few completely different people have to share a flat, very often a number of difficulties emerge.</p> | <p>very often a number (<i>nmod</i>) of difficulties (<i>nmod</i>) emerge (<i>nmod</i>) <i>Вся вторая клауза считается nmod для flat. Проблема возникает из-за того, что emerge распознается как существительное.</i></p> | <p>very often a number (<i>nsubj</i>) of difficulties (<i>nmod</i>) emerge (<i>dobj</i>) <i>Посчитал, что вторая клауза -- прямой объект к share; a flat и very often -- модификаторы к emerge</i></p> | <p>very often a number (<i>dobj</i>) of difficulties (<i>pobj</i>) emerge (<i>root</i> всей клаузы) <i>Ошибка возникает из-за неправильного понимания отношений в клаузе.</i> Нужно посмотреть больше предложений такого типа.</p> |

| | | | |
|--|--|---|---|
| | <p>UDPipe не справился с задачей, поскольку распознал emerge как существительное. Два других парсера поняли, что emerge -- глагол, но тем не менее не смогли правильно понять отношения в предложении.</p> | | |
| <p>VERBS MISTAKEN FOR NOUNS The number of conflicts decreases as the friend or foe dichotomy softens.</p> | <p>as the friend (<i>obl</i>) or foe (<i>amod</i>) dichotomy (<i>nmod</i>) softens (<i>conj</i>) <i>Распознает softens как сказуемое, но считает его noun, из-за этого теряет подлежащее dichotomy.</i></p> | <p>as the friend (<i>nmod</i>) or foe (<i>compound к dichotomy</i>) dichotomy (<i>conj</i>) softens (<i>dep</i>) <i>Посчитал, что friend и dichotomy -- однородные члены, поставил их как модификаторы к root; не понял, что делать с softens</i></p> | <p>as the friend (<i>nsubj</i>) or foe (<i>compound к dichotomy</i>) dichotomy (<i>conj</i>) softens (<i>advcl, зависит от decreases - root</i>) <i>Всё хорошо, не очень хорошо, что принял soften за noun, но синтаксически предложение понято верно</i></p> |
| <p>UDPipe и Spacy приняли глагол 'softens' за существительное. Тем не менее Spacy синтаксически правильно разметил предложение, а единственной ошибкой в синтаксической разметке UDPipe -- неправильное распознавание подлежащего зависимой клаузы.</p> <p>Stanford CoreNLP понял, что 'softens' -- глагол, но разметил предложение хуже, чем остальные парсеры, не поняв, как размечать глагол.</p> | | | |
| <p>They pollution our planet and some ther replicate product give hits for atmosfere.</p> | <p>They (<i>nsubj</i>) pollution (<i>nmod</i>) our planet (<i>nsubj</i>) and some ther replicate product (<i>conj, NOUN</i>) give (<i>root</i>) hits for atmosfere. <i>Верно определил вершину второй клаузы, но не нашел ее подлежащее, посчитав pollution и planet nsubj-ами. С первым и вовсе не справился. Выделил одну лишнюю клаузу для 'and some ther replicate product'</i></p> | <p>They (<i>nsubj</i>) pollution (<i>root, NOUN</i>) our planet (<i>dobj</i>) and some (<i>conj</i>) ther replicate product give hits for atmosfere. <i>Первое сочетание подлежащее -- сказуемое отметил правильно, но вторую клаузу просто прикрепил к прямому объекту и там вообще не разобрать что к чему.</i></p> | <p>They (<i>nsubj</i>) pollution (<i>root, verb</i>) our planet (<i>dobj</i>) and some (<i>det</i>) ther replicate product(<i>conj от planet</i>) give hits for atmosfere. <i>"Они загрязняют нашу планету" - распознано верно, где вершина, где verb по смыслу, вторая часть не распознается так, как хотел бы этого русскоязычный автор, но здесь проблема в абсолютно неправильном предложении, из-за чего получается, что product-однородное прямое дополнение с planet, а не подлежащее во второй клаузе.</i></p> |
| <p>Stanford CoreNLP и Spacy правильно разметили первую клаузу, несмотря на ошибку в написании глагола. Первый маркировал 'pollution' как noun, а второму удалось понять, что это неправильно написанный глагол. UDPipe не справился с первой клаузой из-за ошибки в написании глагола.</p> | | | |

Вторую клаузу парсерам не удалось правильно разметить по причине слишком неграмотного написания. Только UDPipe смог определить вершину второй клаузы.

Table 2.

| Sentence | UDPipe | Stanford CoreNLP | Spacy |
|---|---|--|---|
| <p>NO COPULA Naomy is expected to be nominated as the winner of competition between the models, because she really beautiful. - esl_01293.txt</p> <p>Naomy is expected to be nominated as the winner of competition between the models, because she is really beautiful. - esl_01293.txt</p> | <p>beautiful(<i>advcl</i>), she(<i>nsubj</i>) - <i>прилагательное является вершиной зависимой клаузы</i></p> <p>she(<i>nsubj</i>), is(<i>cop</i>), beautiful(<i>advcl</i>) - <i>прилагательное - вершина зависимой клаузы, копула зависит от прилагательного</i></p> | <p>she(<i>nsubj</i>) really beautiful(<i>advcl</i>)</p> <p><i>correct; marked the adjective as the head of the dependent clause</i></p> <p>she (<i>nsubj</i>) is (<i>cop</i>) really beautiful (<i>advcl</i>)</p> <p><i>correct; marked the adjective as the head of the dependent clause</i></p> | <p>beautiful(<i>advcl</i>), she(<i>nsubj</i>) - <i>прилагательное является вершиной зависимой клаузы</i></p> <p>she(<i>nsubj</i>), is(<i>advcl</i>), beautiful(<i>acomp</i>) - <i>копула - вершина зависимой клаузы, именная часть сказуемого зависит от копулы</i></p> |
| | <p>Absence of copula did not change the syntactic hierarchy of the dependent clause in the trees that were made by UDPipe and Stanford CoreNLP: the predicate adjective is the head of the dependent clause and the copula is connected to it or is absent.</p> <p>As we know, Spacy's parsing logic places the copula as the head of the clause (in a grammatically constructed sentence). In the sentence in which the copula is not absent Spacy expectedly put the copula as the head of the dependent clause and connected the predicate adjective to it. However, in the sentence, in which the copula is missing, the parser put the predicate adjective in the head of the dependent clause; thus, the trees for sentences in which the copula is absent are identical for all three parsers.</p> | | |
| <p>NO COPULA In contrast, other people claim that facilities not so important.</p> <p>In contrast, other people claim that facilities are not so important.</p> | <p>that (<i>mark</i>) facilities (<i>nsubj</i>) not so important (<i>ccomp</i>) <i>вершина important зависит от root claim</i></p> <p>that (<i>mark</i>) facilities (<i>nsubj</i>) are (<i>cop</i>) not so important (<i>ccomp</i>) <i>вершина important зависит от root claim</i></p> | <p>that (<i>doobj</i>) facilities (<i>dep</i>) not so important (<i>advmod</i>) <i>the parser didn't find the dependent clause; marked 'that' as a direct object and connected everything else to it</i></p> <p>that (<i>mark</i>) facilities (<i>nsubj</i>) are (<i>cop</i>) not so important (<i>ccomp</i>) <i>correct; marked the adjective as the head of the dependent clause</i></p> | <p>important(<i>ccomp</i>) - <i>вершина зависимой клаузы, facilities(nsubj)</i></p> <p>are(<i>ccomp</i>) - <i>вершина зависимой клаузы, important(acomp), facilities(nsubj)</i></p> |
| | <p>UDPipe, Stanford Core NLP and UDPipe all correctly parsed the sentence in which the copula is present, each by their own parsing logic.</p> | | |

| | | | |
|--|--|---|---|
| | The absence of the copula did not prevent UDPipe and Spacy from parsing the sentence correctly. In the former's syntactic tree the adjective stayed as the head of the dependent clause, and in the latter's, the adjective moved to the position of head, as it did in the previous pair of sentences. Stanford CoreNLP could not correctly parse the sentence with an absent copula because it did not recognize the dependent clause. | | |
| NO COPULA It is shown in the picture below. It shown in the picture below. | it (<i>nsubjpass</i>) is (<i>auxpass</i>) shown (<i>root</i>) <i>все хорошо</i> | it (<i>nsubjpass</i>) is (<i>auxpass</i>) shown (<i>root</i>) <i>correct</i> | it (<i>nsubjpass</i>), is (<i>auxpass</i>), shown(<i>root, verb</i>) it(<i>nsubj</i>) shown (<i>root, verb</i>) <i>не понял, что пассив, в остальном ок</i> |
| | it (<i>root</i>) shown (<i>advcl</i>) in the picture below. <i>не справился: разделил на две клаузы с нелогичным it-root</i> | it (<i>nsubj</i>) shown (<i>root</i>) <i>correct, except for the fact that the parser did not recognize that the sentence is passive</i> | |
| | All three correctly parsed the sentence in which the copula is present. It is also important to note that Spacy does not put the copula in the head of the passive clause because a semantic verb occupies that place. In the second sentence Stanford CoreNLP and Spacy did not recognize that the construction is passive, but otherwise correctly parsed the sentence and placed the semantic verb as the head of the clause. UDPipe could not correctly parse the sentence with an absent copula. | | |
| PUNCTUATION INFLUENCE Componential analysis, or lexical decomposition, is method of presenting the sense of lexeme through it parts. Componential analysis or lexical decomposition is method of presenting the sense of lexeme through it parts. | ничего не меняется, все распознано верно | Componential analysis (<i>nsubj</i>), or (<i>cc</i>) lexical decomposition (<i>conj</i>), is (<i>cop</i>) method (<i>root</i>) Componential analysis (<i>nsubj</i>) or (<i>cc</i>) lexical decomposition (<i>conj</i>) is (<i>cop</i>) method (<i>root</i>) <i>The parser marked both sentences identically</i> | Componential analysis (<i>nsubj</i>), or (<i>cc</i>) lexical decomposition (<i>conj</i>), is (<i>root-AUX</i>) method (<i>attr npi is-root</i>) Componential analysis (<i>nsubj</i>), or (<i>cc</i>) lexical decomposition (<i>conj</i>), is (<i>root-AUX</i>) method (<i>attr npi is-root</i>) <i>одинаково разметил</i> |
| | | Absence or presence of commas does not influence the parsing of this sentence by all three parsers. | |
| PUNCTUATION INFLUENCE 73% of them prefer facebook, 25% use Instagram and 31% go to LinkedIn. 73% of them prefer facebook, 25% use | 73% (<i>nsubj</i>) of them prefer (<i>root</i>) facebook, 25% (<i>conj</i>) use (<i>mark</i>) Instagram (<i>nsubj</i>) and 31% (<i>conj</i>) go (<i>advcl</i>) to LinkedIn. <i>Первая клауза выделена правильно, вторая и третья -- совершенно нет. Парсер</i> | 73% (<i>nsubj</i>) of them prefer (<i>root</i>) facebook, 25% (<i>amod to Instagram</i>) use (<i>compound for Instagram</i>) Instagram (<i>nsubj to go</i>) and 31% (<i>conj to go</i>) go (<i>parataxis; depends from root</i>) to LinkedIn. | 73% (<i>%-nsubj</i>) of them prefer (<i>ccomp, зависум от use</i>), 25% (<i>% - nsubj от use</i>) use (<i>root</i>) Instagram (<i>dobj от use</i>) and 31% (<i>nsubj к go</i>) go (<i>conj, зависум от root</i>) to LinkedIn. |

| | | | |
|--|---|---|--|
| <p>Instagram, and 31% go to LinkedIn.</p> | <p><i>увидел 4 клаузы, втч просто 31% и просто 25%.</i></p> <p>73% (<i>nsubj</i>) of them prefer (<i>root</i>) facebook, 25% (<i>nsubj</i>) use (<i>conj</i>) Instagram (<i>obj</i>), and 31% (<i>nsubj</i>) go (<i>conj</i>) to LinkedIn.</p> <p><i>В случае с запятой все определено совершенно верно.</i></p> | <p><i>The first clause is marked correctly.</i></p> <p><i>The parser incorrectly recognized '25%' and 'use' as a modifier and a compound to 'Instagram', respectively.</i></p> <p><i>The parser incorrectly recognized the third verb as a verb of a parataxis clause, marked 'Instagram' and '31%' as subjects</i></p> <p>73% (<i>nsubj</i>) of them prefer (<i>root</i>) facebook, 25% (<i>amod to Instagram</i>) use (<i>compound for Instagram</i>) Instagram (<i>dobj to prefer</i>) and 31% (<i>nsubj</i>) go (<i>conj; depends from Instagram</i>) to LinkedIn.</p> <p><i>The first clause is marked correctly.</i></p> <p><i>The parser falsely marked 'Instagram' as a direct object to 'prefer'.</i></p> <p><i>The parser incorrectly recognized '25%' and 'use' as a modifier and a compound to 'Instagram', respectively..</i></p> <p><i>The third clause is marked correctly, but is falsely connected to 'Instagram'; instead it should be connected to root.</i></p> | <p>73% (<i>%-nsubj</i>) of them prefer (<i>root</i>), 25% (<i>%-nsubj om use</i>) use (<i>xcomp,завуcum om prefer-root</i>) Instagram (<i>dobj om use</i>) and 31% (<i>nsubj к go</i>) go (<i>conj, завуcum om prefer-root</i>) to LinkedIn.</p> |
| | <p>Spacy was the best at parsing the sentence that did not have a comma before 'and'. It recognized all three clauses and its only mistake was that it marked the first verb as a clausal complement to the second verb. UDPipe and Stanford CoreNLP correctly handled the first clause, but could not accurately parse the second and the third clauses.</p> <p>The sentence with a comma before 'and' was parsed perfectly by UDPipe: it recognized all three clauses and the syntactic relationships between them. Spacy</p> | | |

| | | | |
|--|--|---|---|
| | made a single mistake: it wrongly recognized the second verb as an open clausal complement. Stanford CoreNLP could not correctly parse the sentence even with the correct punctuation in place because it wrongly recognized some elements as objects and modifiers.. | | |
| <p>PUNCTUATION INFLUENCE</p> <p>As we can see, the most popular place among the foreigners is North America, the number of visitors have increased since 1990 till 2000, and decreased a little bit to 2005.</p> <p>As we can see, the most popular place among the foreigners is North America, the number of visitors have increased since 1990 till 2000 and decreased a little bit to 2005.</p> | <p>ничего не меняется, вторая клауза разделяется и не распознается <i>nsubj</i> place</p> | <p>the most popular place (<i>nsubj</i>) among the foreigners is (<i>cop</i>) North America (<i>root</i>), the number (<i>nsubj</i>) of visitors have (<i>aux</i>) increased (<i>parataxis</i>) since 1990 till 2000, and (<i>cc</i>) decreased (<i>conj</i>) a little bit to 2005.</p> <p><i>'increased' is the verb of a parataxis clause that is connected to root 'decreased' is marked as a homogeneous verb with 'increased'</i></p> <p>the most popular place (<i>nsubj</i>) among the foreigners is (<i>cop</i>) North America (<i>root</i>), the number (<i>nsubj</i>) of visitors have (<i>aux</i>) increased (<i>parataxis</i>) since 1990 till 2000 and (<i>cc</i>) decreased (<i>conj</i>) a little bit to 2005.</p> <p><i>increased' is the verb of a parataxis clause that is connected to root 'decreased' is marked as a homogeneous verb with 'increased'</i></p> | <p>the most popular place (<i>nsubj</i>) among the foreigners is (<i>ccomp om increased-root</i>) North America (<i>attr om is</i>), the number (<i>nsubj om increased-root</i>) of visitors have (<i>aux</i>) increased (<i>root</i>) since 1990 till 2000, and (<i>cc</i>) decreased (<i>conj om increased</i>) a little bit to 2005. - в обоих случаях</p> <p>Наличие запятой не повлияло на правильное понимание структуры всего предложения.</p> |
| | <p>Absence of comma did not interfere with the work of the parsers.</p> <p>Stanford CoreNLP parsed everything correctly, marking the second verb as a parataxis from the first verb and finding the homogeneous verbs in the second clause. Spacy wrongly marked the first verb as a clausal complement to the second verb and UDPipe could not recognize the subject of the first clause.</p> | | |
| <p>There are apples and pineapples on the table.</p> | <p>Все как и ожидалось</p> | <p>Everything is correct. apple(s) - <i>nsubj</i> is/are - <i>root</i></p> | <p>are(<i>root</i>), apples(<i>attr</i>), pineapples(<i>conj</i>, зависит от apples)</p> |

| | | | |
|---|--|--|---|
| <p>There are an apple and pineapples on the table.</p> <p>There is apples and pineapples on the table.</p> <p>There is an apple and pineapples on the table.</p> | | | <p>are(<i>root</i>), apple(<i>attr</i>), pineapples(<i>conj</i>, зависит от apple)</p> <p>is(<i>root</i>), apples(<i>attr</i>), pineapples(<i>conj</i>, зависит от apples)</p> <p>is(<i>root</i>), apple(<i>attr</i>), pineapples(<i>conj</i>, зависит от apples)</p> |
| <p>All sentences were parsed correctly by all three parsers. Absence of an article and incorrect agreement did not disrupt the work of the parsers.</p> | | | |
| <p>VERBS MISTAKEN FOR NOUNS</p> <p>When a few completely different people have to share a flat, very often a number of difficulties emerge.</p> | <p>very often a number (<i>nmod</i>) of difficulties (<i>nmod</i>) emerge (<i>nmod</i>) <i>Вся вторая клауза считается nmod для flat. Проблема возникает из-за того, что emerge распознается как существительное.</i></p> | <p>very often a number (<i>nsubj</i>) of difficulties (<i>nmod</i>) emerge (<i>dobj</i>) <i>The parser falsely recognized the verb of the second clause as a direct object to 'share' and incorrectly marked 'a flat' and 'very often' as modifiers to 'emerge'</i></p> | <p>very often a number (<i>dobj</i>) of difficulties (<i>pobj</i>) emerge (<i>root</i> всей клаузы) <i>Ошибка возникает из-за неправильного понимания отношений в клаузе. Нужно посмотреть больше предложений такого типа.</i></p> |
| <p>UDPipe couldn't correctly parse the sentence because it marked 'emerge' as a noun. Two other parsers accurately marked 'emerge' as a verb, but too failed to correctly parse the sentence as they falsely recognized the syntactic relationships in the sentence.</p> | | | |
| <p>VERBS MISTAKEN FOR NOUNS</p> <p>The number of conflicts decreases as the friend or foe dichotomy softens.</p> | <p>as the friend (<i>obl</i>) or foe (<i>amod</i>) dichotomy (<i>nmod</i>) softens (<i>conj</i>) <i>Распознает softens как сказуемое, но считает его noun, из-за этого теряет подлежащее dichotomy.</i></p> | <p>as the friend (<i>nmod</i>) or foe (<i>compound к dichotomy</i>) dichotomy (<i>conj</i>) softens (<i>dep</i>) <i>Falsely recognized 'friend' and 'dichotomy' as homogeneous nouns, marked them as modifiers to root. The parser could not determine what is 'softens', and marked it as 'dep'.</i></p> | <p>as the friend (<i>nsubj</i>) or foe (<i>compound к dichotomy</i>) dichotomy (<i>conj</i>) softens (<i>advcl</i>, зависит от decreases - <i>root</i>) <i>Всё хорошо, не очень хорошо, что принял soften за noun, но синтаксически предложение понято верно</i></p> |
| <p>UDPipe and Spacy falsely marked 'softens' as a noun. However, the latter otherwise correctly parsed the sentence and the only other mistake that the former did was that it did not recognize the subject of the dependent clause.</p> <p>Stanford CoreNLP correctly marked 'softens' as a verb, but parsed the sentence worse than the other parsers.</p> | | | |
| <p>They pollution our planet and some ther replicate</p> | <p>They (<i>nsubj</i>) pollution (<i>nmod</i>) our planet (<i>nsubj</i>) and some ther</p> | <p>They (<i>nsubj</i>) pollution (<i>root, NOUN</i>) our planet (<i>dobj</i>) and some (<i>conj</i>)</p> | <p>They (<i>nsubj</i>) pollution (<i>root, verb</i>) our planet (<i>dobj</i>) and some (<i>det</i>) ther</p> |

| | | | |
|---|--|--|---|
| <p>product give hits for atmosfere.</p> | <p>replicate product (<i>conj, NOUN</i>) give (<i>root</i>) hits for atmosfere.</p> <p><i>Верно определил вершину второй клаузы, но не нашел ее подлежащее, посчитав pollution и planet nsubj-ами. С первым и вовсе не справился. Выделил одну лишнюю клаузу для 'and some ther replicate product'</i></p> | <p>ther replicate product give hits for atmosfere.</p> <p><i>The parser marked the first noun and the first verb correctly (although it did mark the misspelled verb as a noun).</i></p> <p><i>The parser incorrectly connected the second clause to the direct object 'planet'.</i></p> | <p>replicate product(<i>conj om planet</i>) give hits for atmosfere.</p> <p><i>"Они загрязняют нашу планету" - распознано верно, где вершина, где verb по смыслу, вторая часть не распознается так, как хотел бы этого русскоязычный автор, но здесь проблема в абсолютно неправильном предложении, из-за чего получается, что product-однородное прямое дополнение с planet, а не подлежащее во второй клаузе.</i></p> |
| <p>The first clause was parsed correctly by Stanford CoreNLP and Spacy. The latter additionally understood that the misspelled verb 'pollution' is a verb; the former marked it as a noun. UDPipe could not parse the first clause correctly due to the misspelling of the verb.</p> <p>The second clause couldn't be parsed correctly by any of the parsers. Only UDPipe correctly marked the head of the second clause, albeit it falsely put it as the root of the sentence.</p> | | | |

Приложение 1.

Список примеров для более подробной проверки работы парсеров, результаты можно посмотреть здесь():

Excellent phone to have but unfortunately the unlocked version doesn't have video calling.

The figure demonstrates the popularity of various social networks among different age groups in the USA.

The figure demonstrate the popularity of various social networks among different age groups in the USA.

I strongly believe that a 1,5 year period is enough for an athlete to realize what he has done and not do this ever again.

I strongly believe that a 1,5 year period is enough for an athlete to realize what he have done and not do this ever again.

Artists always have been people which gave us some inspiration.

Artists always has been people which gave us some inspiration.

The figure demonstrates the popularity of various social networks among different age groups in the USA.

The figure demonstrate the popularity of various social networks among different age groups in the USA.

A lot of people lose their study places due to this experiment.

A lot of people loses their study places: due to this experiment.

A lot of polution goes to our atmosphere from planes' oil, because its TSI very dangerous for our world.

A lot of pollution goes to our atmosphere from planes' oil, because it's very dangerous for our world.
When a few completely different people have to share a flat, very often a number of difficulties emerge.
When a few completely different people have to share a flat, very often a number of difficulties emerge.
The number of conflicts decreases as the friend or foe dichotomy softens.
The number of conflicts decrease as the friend or foe dichotomy softens.

The reasons for increased attention to this topic are the jump of the frequency of law breaks among youngers.

The reasons for increased attention to this topic is the jump of the frequency of law breaks among youngers.

Meanwhile, the level of children without access to primary education in South Asia has shown rapid decrease from 2000 to 2012.

Meanwhile, the level of children without access to primary education in South Asia have shown rapid decrease from 2000 to 2012.

Music as well as film industry are complicated business systems made for our entertainment.

Music as well as film industry is a complicated business system made for our entertainment.

To sum up, the length of route and the amount of passengers do not depend on each other.

To sum up, the length of route and the amount of passengers does not depend on each other.

But they, on the other hand, are awful and don't fit cultural norms, accepted in society.

But they, on the other hand, is awful and doesn't fit cultural norms, accepted in society.

But the longer the person lives the more material objects he or she gains.

But the longer the person lives the more material objects he or she gain.

Neither a little kid nor an adult is able to exist comfortable without technics.

Neither a little kid nor an adult are able to exist comfortable without technics.

Neither little kid nor adult is able to exist comfortable without technics.

Neither little kid nor adult are able to exist comfortable without technics.

So personally he or she neither agree nor disagree with the given opinion.

So personally he or she neither agrees nor disagrees with the given opinion.

In my country more than 80 percent of people have it.

In my country more than 80 percent of people has it.

In contrast, only 11% of road transportation consists of metal and food, accordingly.

In contrast, only 11% of road transportation consist of metal and food, accordingly.

73% of them prefer facebook, 25% use Instagram and 31% go to Linked In.

73% of them prefers facebook, 25% uses Instagram and 31% goes to Linked In.

The government calls it "healthy lifestyle popularization".

There are several opinions on this issue that I am going to discuss.

There is several opinions on this issue that I am going to discuss.

There are a lot of examples of occupations where the majority of successful persons is only male : soldier, fireman, builder, ect.

There is a lot of examples of occupations where the majority of successful persons is only male : soldier, fireman, builder, ect.

However, it can't be called an unehiable rule, and of course, there are a plenty of exceptions.

However, it can't be called an unehiable rule, and of course, there is a plenty of exceptions.

Maybe I would have gone to France for several months and would speak French as fluently as the French do.

Maybe I would have gone to France for several months and would speak French as fluently as the French does.

The rich always want more.

The rich always wants more.

The reach always want more.

The reach always wants more.

Having thoughts of committing crimes comes from the period of socialization.

Having thoughts of committing crimes come from the period of socialization.

So "Nine stories" is a collection of nine short stories.

So "Nine stories" are a collection of nine short stories.

The bachelor's degree have a 45 per cent of males and 55 per cent of females.

But in the same time the postgraduate diploma have only 30%, while these number among the men is accounted for 70%.

The differences there is quite big, for instance skilled vocational diploma have 10% of women and 90% of men.

It is I who am responsible.

It is I who is responsible.

It is the secretary of our bosses who sends the bills to Mr. Harding.

It is the secretary of our bosses who send the bills to Mr. Harding.

Приложение 1.

Список примеров для второй проверки работы парсеров, результаты можно посмотреть [здесь\(\)](#):

Naomy is expected to be nominated as the winner of competition between the models, because she really beautiful. - esl_01293.txt

Naomy is expected to be nominated as the winner of competition between the models, because she is really beautiful. - esl_01293.txt

In contrast, other people claim that facilities not so important.

In contrast, other people claim that facilities are not so important.

It is shown in the picture below.

It shown in the picture below.

Componential analysis, or lexical decomposition, is method of presenting the sense of lexeme through it parts.

Componential analysis or lexical decomposition is method of presenting the sense of lexeme through it parts.

73% of them prefer facebook, 25% use Instagram and 31% go to LinkedIn.

73% of them prefer facebook, 25% use Instagram, and 31% go to LinkedIn.

As we can see, the most popular place among the foreigners is North America, the number of visitors have increased since 1990 till 2000, and decreased a little bit to 2005.

As we can see, the most popular place among the foreigners is North America, the number of visitors have increased since 1990 till 2000 and decreased a little bit to 2005.

There are apples and pineapples on the table.

There are an apple and pineapples on the table.

There is apples and pineapples on the table.

There is an apple and pineapples on the table.

When a few completely different people have to share a flat, very often a number of difficulties emerge.

The number of conflicts decreases as the friend or foe dichotomy softens.

Music as well as film industry is a complicated business system made for our entertainment.

They pollution our planet and some ther replicate product give hits for atmosfere.