



Разметка ошибок в учебных текстах на немецком языке

Солнышков А. А. и Ремшу Н. Л.
(Петрозаводский государственный университет)



Учебные корпуса

Корпус учебных текстов – это электронное собрание структурированных, чаще всего аннотированных текстов (сочинений, описаний, писем, тестов и т.д.) обучающихся иностранному языку.



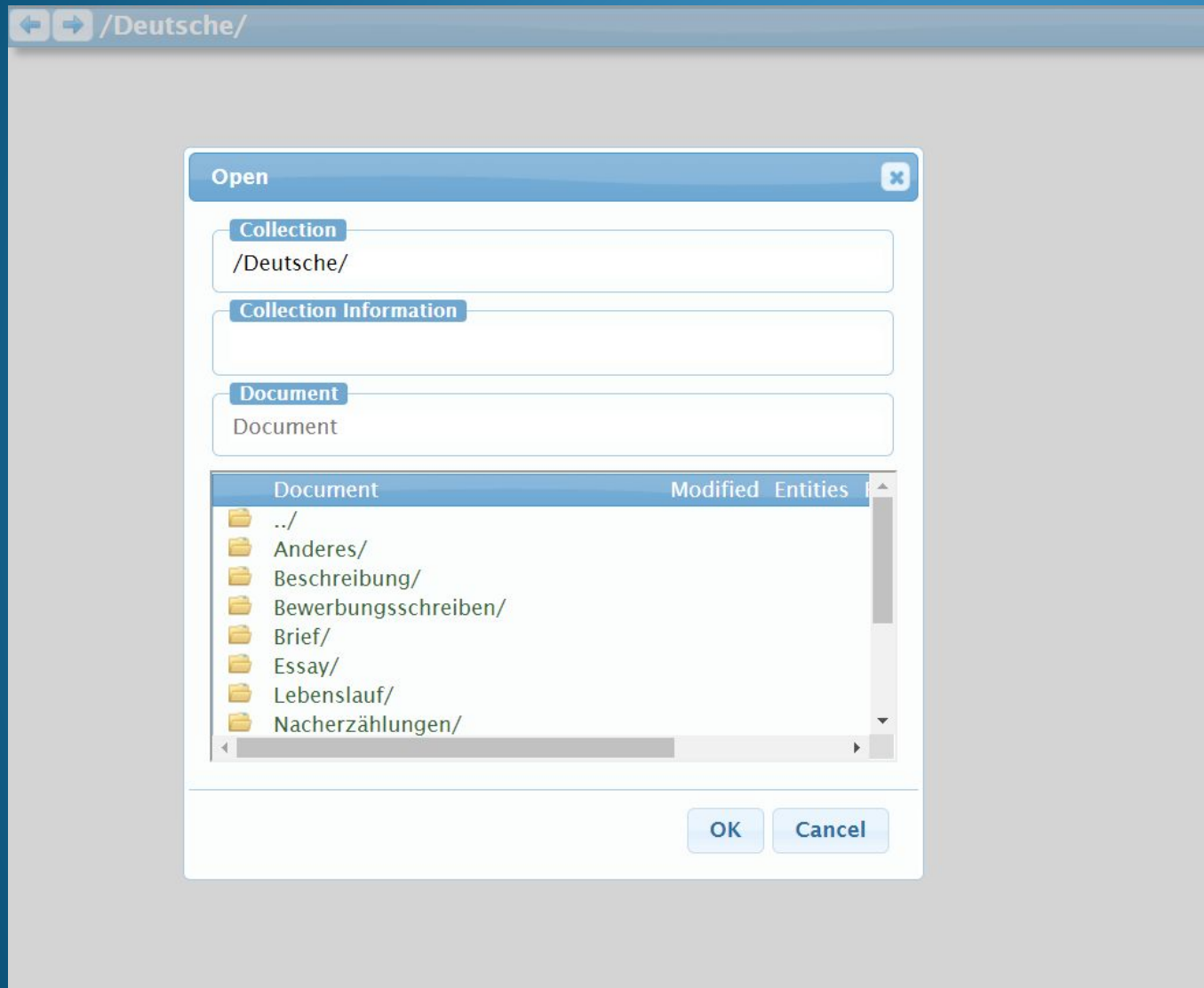
Цель

При разметке ошибок передавать студенту информацию о

- сути этой ошибки
- вариантах ее исправления
- возможной причине ее возникновения
- степени влияния этой ошибки на понимание высказывания

Также дать преподавателю поле для исследования работ студентов одного или разных курсов.

Система разметки учебного корпуса



Brat Rapid Annotation Tool - система для аннотирования в корпусе, в нашем случае используется для разметки ошибок

Процесс регистрации студента



Студент должен внести информацию о себе при регистрации для возможных дальнейших исследований

Sign Up

Фамилия(Иванов)

Фамилия

Имя(Иван)

Имя

Отчество(Иванович)

Отчество

Логин

Логин

Пароль

Пароль

Пол

Возраст

Возраст

Курс

Номер группы

Номер группы

Год изучения английского

Год изучения английского

Год изучения французского

Год изучения французского

Год изучения немецкого

Год изучения немецкого

OK

Cancel



Уровни аннотирования ошибки в REALES :

- лингвистическое описание;
- причина ошибки («Интерференция с L1», «Отсутствие данной категории в L1», «Опечатка» или «Прочее»);
- степень критичности для понимания смысла (по 3 ступеням);
- вариант исправления.

Классификация типов ошибок в REALES (отры



- Punctuation
- Spelling
- Capitalisation
- Grammar

Determiners

Articles

Verbs

Tense

Choice of tense

Tense: form

Voice

Modals

Verb pattern

Intransitive verb (sleep)

Transitive verb (eat meat)

Reflexive verb (wash oneself)

Verb with as (regard as, consider as)

Ambitransitive verb (the door opened/John opened the door/The door was opened by ...; The waiter will seat you/The theatre seats 500 people/We were seated in the corner)

Two verbal forms in the predicate

Verb+Infinitive (want to do smth)

Verb+Gerund (avoid doing)

Verb+Infinitive OR Gerund

Verb+Bare Infinitive (dare do)

Verb+Object/Addressee + Bare Infinitive (let smb do, make smb do)

Infinitive Restoration Alternation (was made to do it)

Verb+Participle (be done with)

Get+Participle (get started)



Уровни аннотирования ошибки в FALCO :

- лингвистическое описание;
- определение формального типа ошибки (• добавление • замена • пропуск • перестановка);
- определение «носителя ошибки» (• слог • слово • предложение • фраза • ... имя • глагол • прилагательное ...);
- варианты исправления.

Варианты исправлений в FALCO :

3 ⓘ ↶ Path: FalkoEssayL1v2.0 > dcs015_2007_10_L1v2.0 (tokens 482 - 492)

Eine wirkliche Veränderung der Probleme können nur in einem Disput erfolgen
 ein wirklich Veränderung d Problem können nur in ein Disput erfolgen
 ART ADJA NN ART NN VMFIN ADV APPR ART NN VVIN

⊕ ZHverb (grid)
 ⊕ falko (grid)
 ⊕ learner (grid)
 ⊖ ZH1 (grid)

ZH1	Eine	wirkliche	Veränderung	der	Probleme	kann	nur	in	einem	Disput	erfolgen
ZH1Diff						CHA					
ZH1lemma	ein	wirklich	Veränderung	d	Problem	können	nur	in	ein	Disput	erfolgen
ZH1pos	ART	ADJA	NN	ART	NN	VMFIN	ADV	APPR	ART	NN	VVIN
tok	Eine	wirkliche	Veränderung	der	Probleme	können	nur	in	einem	Disput	erfolgen

⊕ ZH2 (grid)



Аннотирование ошибок в разрабатываемом корпусе ПетрГУ:

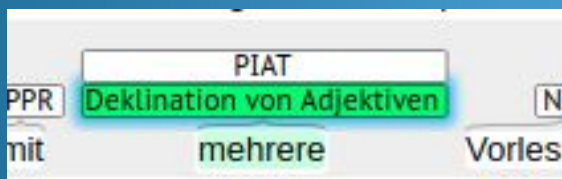
- лингвистическая характеристика ошибки;
- область ошибки (слово, словосочетание, предложение);
- степень критичности ошибки для понимания смысла;
- возможная причина ошибки.

Область ошибки

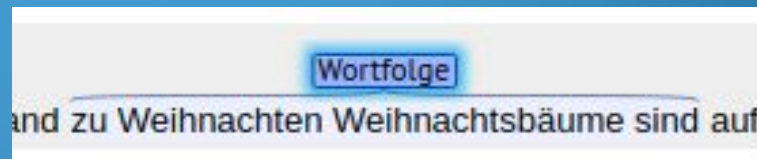
Межсимвольное пространство:



Слово:



Предложение или набор слов:



Степень критичности ошибки для понимания смысла

Auswirkung des Fehlers:	
	beeinflusst_das_Verstehen_nicht
	stört_das_Verstehen
	der_Sinn_ist_unverständlich/verdreht

beeinflusst das Verstehen nicht - не
влияет на понимание

stört das Verstehen - мешает
пониманию

der Sinn ist unverständlich/verdreht -
смысл непонятен/искажен

Возможная причина ошибки

Grund des Fehlers:

- Interferenz_mit_der_Muttersprache
- Interferenz_mit_Englisch
- Tippfehler

Interferenz mit der Muttersprache -
интерференция с родным языком

Interferenz mit Englisch -
интерференция с английским
языком

Tippfehler - опечатка

Классификация типов ошибок в разрабатываемом корпусе ПетрГУ



- ☐ Fehler
 - ☐ Grammatik
 - ☐ Substantive
 - ☐ Geschlecht
 - ☐ Numerus
 - ☐ Deklination
 - ☐ Rektion von Substantiven
 - ☐ Artikel
 - ☐ Bestimmter Artikel
 - ☐ Unbestimmter Artikel
 - ☐ Negationsartikel
 - ☐ Nullartikel
 - ☐ Zahlwörter
 - ☐ Kardinalzahlen
 - ☐ Ordinalzahlen
 - ☐ Weitere Zahlwörter
 - ☐ Pronomen
 - ☐ Personalpronomen
 - ☐ Possessivpronomen
 - ☐ Demonstrativpronomen
 - ☐ Interrogativpronomen
 - ☐ Reflexivpronomen
 - ☐ Deklination des Pronomens

- ☐ Fehler
 - ☐ Грамматика
 - ☐ Существительное
 - ☐ Род
 - ☐ Число
 - ☐ Склонение
 - ☐ Управление существительного
 - ☐ Артикль
 - ☐ Определенный артикль
 - ☐ Неопределенный артикль
 - ☐ Отрицательный артикль
 - ☐ Нулевой артикль
 - ☐ Числительное
 - ☐ Количественное числительное
 - ☐ Порядковое числительное
 - ☐ Другие указатели на число
 - ☐ Местоимение
 - ☐ Личное местоимение
 - ☐ Притяжательное местоимение
 - ☐ Указательное местоимение
 - ☐ Вопросительное местоимение
 - ☐ Возвратное местоимение
 - ☐ Склонение местоимений

Классификация типов ошибок в разрабатываемом корпусе ПетрГУ



- ▣ ○ Verben
 - ▣ ○ Konjugation
 - Modalverben
 - Starke Verben
 - Trennbare Verben
 - Untrennbare Verben
 - ▣ ○ Zeitform
 - Wahl der Zeitform
 - Bildung der Zeitform
 - ▣ ○ Modus
 - Imperativ
 - Konjunktiv
 - ▣ ○ Genus
 - Aktive
 - Passive
 - Zustandspassiv
 - Rektion von Verben
- ▣ ○ Partizipien
 - Partizip I
 - Partizip II
- ▣ ○ Präpositionen
 - Wahl der Präposition
 - Präpositionen mit einem bestimmten Kasus
 - Wechselpräpositionen
- Konjunktionen

- ▣ ○ Глагол
 - ▣ ○ Сопряжение
 - Модальный глагол
 - Сильный глагол
 - Глагол с отделяемой приставкой
 - Глагол с неотделяемой приставкой
 - ▣ ○ Время
 - Выбор временной формы
 - Образование временной формы
 - ▣ ○ Наклонение
 - Императив
 - Конъюнктив
 - ▣ ○ Залог
 - Активный залог
 - Пассивный залог
 - Пассив состояния
 - Управление глаголов
- ▣ ○ Причастие
 - Причастие I
 - Причастие II
- ▣ ○ Предлоги
 - Выбор предлога
 - Предлог с определенным падежом
 - Предлог, управляющий несколькими падежами
- Союзы

Классификация типов ошибок в разрабатываемом корпусе ПетрГУ



- ▣ ○ Adjektive
 - Deklination von Adjektiven
 - Komparativform
 - Superlativform
 - Rektion von Adjektiven
- Adverbien
- Grammatische Kongruenz
- ▣ ○ Wortfolge
 - ▣ ○ Verbstellung
 - Direkte Wortfolge
 - Indirekte Wortfolge
 - Satzklammer
 - Wortfolge in einer Satzreihe
 - ▣ ○ Wortstellung
 - Wortstellung in der Negation
 - Thema-Rhema-Gliederung
- Vergleichskonstruktionen
- ▣ ○ Infinitivkonstruktionen
 - Infinitivkonstruktionen mit zu
 - Infinitivkonstruktionen ohne zu

- ▣ ○ Прилагательное
 - Склонение прилагательного
 - Сравнительная степень
 - Превосходная степень
 - Управление прилагательного
- Наречия
- Грамматическое согласование
- ▣ ○ Порядок слов
 - ▣ ○ Место глагола
 - Прямой порядок слов
 - Обратный порядок слов
 - Рамочная конструкция
 - Порядок слов в придаточном предложении
 - ▣ ○ Место второстепенных членов предложения
 - Порядок слов при отрицании
 - Тема-рематическое членение предложения
- Сравнительные конструкции
- ▣ ○ Инфинитивные конструкции
 - Инфинитивные конструкции с zu
 - Инфинитивные конструкции без zu

Классификация типов ошибок в разрабатываемом корпусе ПетрГУ



- ☐ ○ Lexik
 - Lexemauswahl
 - Feste Wendungen
 - ☐ ○ Wortbildung
 - Derivative Suffixe
 - Derivative Präfixe
 - Zusammengesetzte Wörter
- ☐ ○ Diskurs
 - ☐ ○ Logik
 - Konnektoren
 - Kohärenz
 - Referenz
 - Sprachstil
- Auslassungen
- Überflüssige Elemente
- Orthographie
- Interpunktion

- ☐ ○ Лексика
 - Выбор лексеммы
 - Устойчивые обороты
 - ☐ ○ Словообразование
 - Словообразовательные суффиксы
 - Словообразовательные префиксы
 - Сложные слова
- ☐ ○ Дискурс
 - ☐ ○ Логика
 - Соединительные элементы
 - Связность, последовательность
 - Референтные связи внутри текста
 - Стиль
- Пропуски
- Лишние элементы
- Орфография
- Пунктуация

Хранение разметки



```
T1 Überflüssige_Elemente 83 85 du
A1 Auswirkung_des_Fehlers T1 stört_das_Verstehen
T2 Konjugation 78 82 Hast
#1 AnnotatorNotes T2 Haben
T3 Genus 307 310 das
A2 Auswirkung_des_Fehlers T3 stört_das_Verstehen
#2 AnnotatorNotes T3 der
T4 Orthographie 399 403 muss
A3 Auswirkung_des_Fehlers T4 beeinflusst_das_Verstehen_nicht
#3 AnnotatorNotes T4 Muss
T5 Orthographie 500 503 Sie
A4 Grund_des_Fehlers T5 Tippfehler
A5 Auswirkung_des_Fehlers T5 stört_das_Verstehen
#4 AnnotatorNotes T5 sie
T6 Genus 530 533 Das
A6 Auswirkung_des_Fehlers T6 stört_das_Verstehen
#5 AnnotatorNotes T6 Der
T7 Personalpronomen 816 819 uns
#6 AnnotatorNotes T7 mich
T8 bestimmter_Artikel 1058 1063 einem
A7 Auswirkung_des_Fehlers T8 stört_das_Verstehen
#7 AnnotatorNotes T8 dem
```

Преимущества:

- Легко выгрузить данные для исследования или переноса системы
- В наглядной форме можно получить представление о разметке каждого файла

Недостатки:

- Из-за накопления большого количества файлов, система начинает тратить больше времени на обработку работ

1 Ich und meine Familie.

2 Mein name ist Lisa und ich wohne in Petrosawodsk mit meiner Familie.

3 Meine Eltern heißen Elena und Andrew.

4 Mein Vater ist 43 und mein Mutter ist 42 Jahre alt.

5 Sie sind Ärzte von Beruf und arbeiten zusammen.

6 Ich habe einen kleinen Bruder.

7 Er heißt Sergey und er ist 10 Jahre alt.

8 Er besucht die Schule und lernt sehr gut.

9 Mein Vater hat eine Schwester.

10 Sie heißt Anna und sie wohnt seit fünf Jahren in Finland.

11 Sie ist nicht verheiratet und hat keine Kinder.

12 Wir sehen ihr sehr selten.

13 Sie kommt fast nicht nach Russland.

14 Ich habe auch eine Großmutter.

15 Ihr name ist Olga und sie ist rentnerin.

16 Sie ist Mathematiklehrerin von Beruf.

17 Unsere Familie ist sehr freundlich.

18 Ich liebe meine Familie.

1 Kinder im Internet

2 Heute können sich die Menschen das Leben ohne Internet nicht mehr vorstellen.

3 Wir selbst bemerken nicht, wie das Internet uns verändert.

4 Und es betrifft Kinder noch mehr als Erwachsene.

5 Und ist es nützlich oder gefährlich für Kinder das Internet zu benutzen?

6 Dieses Thema ist sehr aktuell und wichtig.

7 Die Kinder verbringen die meiste freie Zeit im Internet.

8 Sie lernen hier Freunde kennen, kommunizieren und streiten.

9 Virtuelle Realität ist für sie genauso wichtig wie das normale Leben.

10 Leider hat das Internet auch unerwünschte und gefährliche Inhalte.

11 Wenn man sehr lange am Computer sitzt, bekommt man die Probleme mit der Gesundheit.

12 Kinder können auch vom Internet abhängig sein und wenig spazieren.

13 Aber natürlich schließen alle diese Nachteile nicht die Vorteile aus.

14 Im Internet kann man sehr viele nützliche und interessante Information finden.

15 Dort gibt es auch viele Programme, die den Speicher entwickeln.

16 Die Kinder können auch mit den Freunden chatten uns skypen, wenn sie Hilfe bei den Hausaufgaben brauchen.

17 Zum Schluß möchte ich sagen, dass die Zahl der Kinder, die das Internet nutzen, wächst jedes Jahr.

18 In diesem Fall erhalten die meisten von ihnen die Information ohne Aufsicht von Erwachsenen.

19 Die Folgen dieses Gesetzes sind unvorhersehbar.

20 Die Eltern müssen nur mehr auf das Kind achten und so viel Zeit wie möglich mit ihm verbringen.

Lexemauswahl
Auswirkung des Fehlers: der_Sinn_ist_unverständlich/verdreht
"Speicher"
Richtig: Gedächtnis

Список источников



- Грудева Е.В., Бучилова И.А., Волкова Н.А. Корпусы ошибок: целевая аудитория, возможная архитектура корпуса // Вестник Череповецкого государственного университета. 2018. №5(86). С. 63–72.
- Introduction into the basic functionality of the brat rapid annotation tool. URL: <https://brat.nlplab.org/manual.html>
- Камшилова О.Н., Трефилова Г.С. Шаблоны для автоматического поиска ошибок в именных и глагольных группах (анализ корпуса ученических текстов) // URL: <https://studylib.ru/doc/1917875/recommendations-on-paper-formatting>
- КРУТ, Корпус русских учебных текстов // URL: http://web-corpora.net/CoRST/search/index.php?interface_language=ru
- Johns, Tim (1991). Should you be persuaded: Two examples of data-driven learning // URL: https://moodle.ils.uw.edu.pl/pluginfile.php/10011/mod_resource/content/2/Tim%20Johns%20and%20DDL.pdf



Спасибо за внимание!