

Кашин Глеб Игоревич

студент

Ведерников Фёдор Яковлевич

студент

Панин Матвей Евгеньевич

студент

Ожигов Егор Сергеевич

студент

Закон Марк Михайлович

студент

Научный руководитель

Новикова Алена Романовна

магистр, старший преподаватель

ФГБОУ ВО «МИРЭА – Российский технологический университет»

г. Москва

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ ДЛЯ ВЫЯВЛЕНИЯ ТИПИЧНЫХ ОШИБОК СТУДЕНТОВ ПРИ ИЗУЧЕНИИ ТЕХНИЧЕСКОГО АНГЛИЙСКОГО ЯЗЫКА

***Аннотация:** разработана процедура выявления типичных языковых ошибок для подготовки заданий по техническому английскому языку с применением частотного анализа и ассоциативных правил. На материале 300 текстов Write & Improve проанализированы 6 532 правки. Построен частотный профиль ошибок и выявлены их сочетания. На этой основе предложены задания на описание компонентов информационных систем и порядок подбора упражнений преподавателем. Применение процедуры к техническому английскому языку проиллюстрировано модельными примерами, предусматривающими проверку грамматической правильности и точности передачи профессионального содержания.*

Ключевые слова: *технический английский язык, анализ ошибок, обработка естественного языка, интеллектуальный анализ данных, ассоциативные правила, обратная связь.*

Изучение технического английского языка предполагает умение точно описывать работу программ и систем. Владения терминами `server` и `request` для этого недостаточно: ошибки в употреблении артиклей, предлогов и глагольных форм затрудняют понимание. Анализ повторяющихся ошибок позволяет обоснованно подбирать профессионально ориентированные упражнения.

Цифровые средства обучения поддерживают индивидуальную работу над языковыми трудностями [1; 2]. Цель исследования – разработать процедуру выявления типичных ошибок для подготовки заданий по техническому английскому языку. Процедура объединяет классификацию правок, анализ их частоты и совместной встречаемости с подбором упражнений.

Обнаружение ошибки, определение её типа и исправление являются самостоятельными задачами [4]. Инструмент ERRANT сопоставляет исходное и исправленное предложения, выделяет правки и классифицирует их по языковым признакам [3]. Такая классификация позволяет обобщить отдельные исправления и составить профиль языковых трудностей.

Материал и методика анализа. Для анализа использованы файлы `A.dev`, `B.dev` и `C.dev` корпуса W&I+LOCNESS версии 2.1 для BEA-2019 [5]. В выборку вошли 130, 100 и 70 текстов *Write & Improve* из групп A, B и C соответственно, отражающих разные уровни владения английским языком. Обработаны 3 396 предложений и 63 856 токенов – единиц текста, включая знаки препинания. Тексты носителей языка исключены из выборки. Исправления внесены специалистами, а типы правок определены с помощью ERRANT. Лицензия допускает использование корпуса в некоммерческих исследованиях при условии ссылки на описание *Write & Improve* [8].

Частотный профиль построен на материале общезыковых учебных текстов; применение процедуры к техническому английскому языку показано на модельных примерах. Ориентиром при разработке профессионально ориенти-

рованных заданий служит корпус NTUCLE, содержащий проектные предложения студентов инженерных направлений [7].

С использованием Python обработаны предложения и правки аннотатора с идентификатором 0 из файлов M2. Служебные записи исключены. Правки, обозначающие добавление M, удаление U и замену R, сгруппированы по типам ошибок: например, M:DET и R:DET отнесены к категории DET. Категории VERB:FORM и VERB:TENSE сохранены отдельно. Одна правка может охватывать несколько слов.

Для каждого предложения сформировано множество типов ошибок без повторов. Проанализированы пары из 24 типов ошибок; в расчётах учтены также предложения без правок. Для выявления связей между типами ошибок использованы ассоциативные правила. Поддержка определяется как доля предложений, содержащих оба типа ошибок: $P(X \cap Y)$. Достоверность правила $X \rightarrow Y$ – доля предложений с ошибками типа Y среди предложений с ошибками типа X: $P(Y|X)$. Коэффициент подъёма $Lift = P(X \cap Y)/(P(X)P(Y))$ характеризует отношение наблюдаемой частоты совместного появления ошибок к ожидаемой при их независимости. Для отбора повторяющихся сочетаний установлен порог поддержки 1%.

Результаты анализа. Выделены 6 532 правки в 2 312 предложениях. В таблице 1 представлены пять наиболее частых типов исправлений. Категория DET охватывает ошибки в употреблении определителей – артиклей и других слов, уточняющих существительное. Категория OTHER включает разнородные правки, требующие разбора конкретных примеров.

Таблица 1

Наиболее частые типы исправлений

Категория	Число правок	Доля, %
PUNCT – пунктуация	1124	17,21
OTHER – прочие правки	881	13,49
DET – определители	758	11,60
PREP – предлоги	675	10,33
VERB:TENSE – время глагола	458	7,01
Остальные 19 категорий	2636	40,36
Всего	6532	100,00

На 100 исходных токенов приходится 15,77 правки в группе А, 10,53 в группе В и 5,07 в группе С. Нормирование на одинаковый объём текста позволяет сопоставлять частоту правок в группах. Этот показатель характеризует частоту правок, а не долю слов с ошибками. Уровень владения языком учитывается при выборе сложности заданий.

Ошибки формы глагола VERB:FORM и глагольного времени VERB:TENSE совместно встречаются в 54 предложениях: поддержка составляет 1,59%, lift – 2,44. Ошибки числа существительного NOUN:NUM и употребления определителей DET сочетаются в 83 предложениях: поддержка составляет 2,44%, lift – 2,24. Из 200 предложений с ошибками NOUN:NUM в 83 также присутствуют ошибки DET; достоверность правила NOUN:NUM → DET равна 41,5%. Это сочетание служит основанием для совместной отработки числа существительного и употребления определителей.

Ошибки употребления определителей DET и предлогов PREP совместно встречаются в 168 предложениях: поддержка составляет 4,95%, lift – 1,62. Ассоциативный анализ выявляет связи между языковыми трудностями, которые не отражаются при раздельном подсчёте ошибок. Выявленные сочетания определяют тематику комплексных заданий. При интерпретации результатов учитываются уровень владения языком, тема и длина предложения.

Применение к техническому английскому языку. Значение технического термина определяется контекстом [6]. В таблице 2 представлены модельные примеры описания программных систем. Они показывают, как использовать выбранные категории ошибок при подготовке профессионально ориентированных заданий и дополнить проверку грамматической правильности оценкой точности технического содержания.

Таблица 2

Модельные примеры для профессионального контекста

Категория	Пример и исправление
DET и NOUN:NUM	a requests → a request (один запрос)
PREP	depends from the server → depends on the server
VERB:SVA	The server return a value. → The server returns a value.
Терминология	physical memory → virtual memory (если описывается виртуальная)

Для сочетания NOUN:NUM и DET разработан вариант задания на описание компонентов системы. Студент определяет, идёт ли речь об одном объекте, нескольких объектах или классе объектов, затем согласует число существительного и определитель. Для отработки PREP предлагается описание зависимостей между компонентами, а для VERB:SVA – описание действий сервера.

Работа преподавателя включает проверку черновиков, подтверждение правок и составление сводки ошибок. Частотный профиль позволяет определить приоритетные языковые темы, а сочетания ошибок – содержание комплексных упражнений. Терминологические ошибки анализируются с учётом смысла высказывания: редкая подмена понятия может быть существеннее частой пунктуационной ошибки.

В учебной группе предлагается собирать обезличенные работы, выполненные по единым техническим заданиям. Преподаватель проверяет тип правки и сохранение смысла, после чего подбирает упражнения на основе сводки ошибок. Повторные работы позволяют оценивать изменения по тем же категориям. Такая организация работы разграничивает выявление языковых трудностей и оценку результатов обучения.

Заключение. Разработанная процедура объединяет автоматическую классификацию правок и интеллектуальный анализ данных для выявления типичных языковых трудностей. На материале 300 текстов построен частотный профиль и выявлены сочетания ошибок в употреблении определителей, числа существительного, формы и времени глагола. На этой основе предложены задания по техническому английскому языку. Процедура определяет порядок подбора упражнений преподавателем с учётом частоты и совместной встречаемости ошибок и предусматривает проверку точности передачи профессионального содержания.

Список литературы

1. Соколов, А. Д. Технические средства обучения иностранному языку / А. Д. Соколов // Образование, инновации, исследования как ресурс развития общества. – Чебоксары : Среда, 2026. – DOI 10.31483/r-155887.
2. Элизбарян, А. Г. Использование современных технологий и искусственного интеллекта в обучении английскому языку / А. Г. Элизбарян, А. С. Климаш, И. Ш. Каппушева // Образование, инновации, исследования как ресурс развития сообщества. – Чебоксары : Среда, 2026. – URL: <https://phsreda.com/e-articles/10872/Action10872-155913.pdf> (дата обращения: 10.10.2026). – Текст : электронный.
3. Bryant, C. Automatic Annotation and Evaluation of Error Types for Grammatical Error Correction / C. Bryant, M. Felice, T. Briscoe // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. – 2017. – Vol. 1. – P. 793–805. – DOI 10.18653/v1/P17-1074.
4. Grammatical Error Correction: A Survey of the State of the Art / C. Bryant, Z. Yuan, M.R. Qorib [et al.] // Computational Linguistics. – 2023. – Vol. 49, No. 3. – P. 643–701. – DOI 10.1162/coli_a_00478.
5. The BEA-2019 Shared Task on Grammatical Error Correction / C. Bryant, M. Felice, Ø.E. Andersen, T. Briscoe // Proceedings of the Fourteenth Workshop on Innovative Use of NLP for Building Educational Applications. – ACL, 2019. – P. 52–75. – DOI 10.18653/v1/W19-4406.
6. Chung, T. M. Technical vocabulary in specialised texts / T. M. Chung, P. Nation // Reading in a Foreign Language. – 2003. – Vol. 15, No. 2. – P. 103–116. – DOI 10.64152/10125/66770.
7. NTUCLE: Developing a Corpus of Learner English to Provide Writing Support for Engineering Students / R.V.P. Winder, J. MacKinnon, S.Y. Li [et al.] // Proceedings of the 4th Workshop on Natural Language Processing Techniques for Educational Applications. – AFNLP, 2017. – C. 1–11. – URL: <https://aclanthology.org/W17-5901/> (дата обращения: 10.10.2026). – Текст : электронный.

8. Developing an automated writing placement system for ESL learners / H. Yannakoudakis, Ø.E. Andersen, A. Geranpayeh [et al.] // Applied Measurement in Education. – 2018. – Vol. 31, No. 3. – P. 251–267. – DOI 10.1080/08957347.2018.1464447.