

ЛИНГВИСТИЧЕСКОЕ ОБРАЗОВАНИЕ

Научная статья

УДК 372.881.111.22

<https://doi.org/10.23951/2307-6127-2024-1-54-64>

Лонгитюдные исследования языковых ошибок на базе немецкоязычного корпуса студенческих текстов

Ирина Аврамовна Котюрова¹, Георгий Романович Сафонов²

^{1,2} *Петрозаводский государственный университет, Петрозаводск, Россия*

¹ *koturova@petrsu.ru*

² *jiri.safonov@gmail.com*

Аннотация

В эпоху цифровизации и активного распространения корпусных технологий в лингвистическом образовании специалисты в области лингводидактики постоянно открывают все новые возможности в работе с большими данными. Одним из относительно новых явлений в российском образовании является сбор корпусов ученических текстов на иностранном языке. Их возможности для лингводидактических исследований зависят прежде всего от продолжительности сбора данных и от разметки, которую содержит тот или иной корпус. В статье речь идет о корпусе немецкоязычных студенческих текстов Петрозаводского аннотированного корпуса текстов и лонгитюдном исследовании типов лингвистических ошибок, допускаемых студентами на всем протяжении изучения немецкого языка в течение 5 лет. Результатом исследования стала статистика по 90 классам ошибок, поделенных на семь основных групп, и динамика этой статистики на протяжении 5 лет обучения немецкому языку. Сравнение наиболее частотных ошибок на первом и пятом курсах показывает, что к темам, которые на протяжении всех годов обучения вызывают наибольшие проблемы у студентов, относятся выбор лексики, орфография, пропуски, пунктуация и обратный порядок слов. Уходя к концу обучения в вузе проблемы с неопределенным артиклем, склонением прилагательных и существительных, образованием множественного числа и родом существительных, уступая место другим проблемам, таким как лишние элементы, логика, порядок слов в придаточных предложениях и стилистические ошибки.

Ключевые слова: *ученический корпус, немецкий язык как иностранный, языковые ошибки, анализ образовательных данных*

Для цитирования: Котюрова И. А., Сафонов Г. Р. Лонгитюдные исследования языковых ошибок на базе немецкоязычного корпуса студенческих текстов // Научно-педагогическое обозрение. 2024. Вып. 1 (53). С. 54–64. <https://doi.org/10.23951/2307-6127-2024-1-54-64>

LINGUISTIC EDUCATION

Original article

Longitudinal studies of language errors based on a German-language learner corpus

Irina A. Kotyurova¹, Georgiy R. Safonov²

^{1,2} *Petrozavodsk State University, Petrozavodsk, Russian Federation*

¹ *koturova@petrsu.ru*

² *jiri.safonov@gmail.com*

© И. А. Котюрова, Г. Р. Сафонов, 2023

Abstract

In the age of digitalization and the active spread of corpus technologies in linguistic education, linguodidactics specialists are constantly discovering new opportunities in working with big data. One relatively new phenomenon in Russian education is the collection of corpora of student texts in a foreign language. It's possibilities for linguodidactical research depend primarily on the duration of the data collection and on the markup that corpus contains. The article focuses on the corpus of German-language student texts PACT (Petrozavodsk annotated corpus of texts) and longitudinal research of types of linguistic mistakes made by students throughout the study of the German language for 5 years. The result of the research is statistics for 90 classes of errors, divided into 7 major groups – grammar, vocabulary, orthography, punctuation, discourse, omissions and superfluous elements – and the dynamics of these statistics over the 5 years of German language study. Comparison of the most frequent errors made by 1st and 5th year students respectively shows that subjects causing the most problems for students during all years of study are lexeme selection, orthography, omissions in text, punctuation and reverse word order. At the end of study problems with indefinite articles, adjective and noun declension, formation of plural form and gender of nouns are giving way to other issues such as superfluous elements in text, logic, word order in subordinate sentences and stylistic errors.

Keywords: learner corpus, German as a foreign language, language errors, educational data mining

For citation: Kotyurova I. A., Safonov G. R. Longitudinal studies of language errors based on a German-language learner corpus. *Nauchno-pedagogicheskoye obozreniye – Pedagogical Review*, 2024, vol. 1 (53), pp. 54–64. <https://doi.org/10.23951/2307-6127-2024-1-54-64>

Введение

Цифровизация, определенная на федеральном уровне как одно из приоритетных направлений развития, активно развивается во всех сферах как повседневной, так и научной жизни. Не являются исключениями методы исследований и образовательные технологии, где большие данные помогают в вопросах обучения, управления образованием и автоматизации оценки работы учащихся [1–4].

Сбор таких электронных баз данных, называемых в лингвистике корпусами, т. е. коллекций аутентичных¹ текстов, позволяет автоматизировать ручной поиск и объективно анализировать большие объемы языковых данных, открывая тем самым невиданные в «докорпусную» эпоху возможности для исследований. Одной из разновидностей лингвистических корпусов является ученический корпус – learner corpus, о пользе которого в преподавании иностранных языков одной из первых заговорила С. Грейнджер [5]. Ученические корпусы – это большие компьютеризированные коллекции письменных и/или устных текстов, которые были созданы учащимися, изучающими иностранный язык, и которые могут быть проанализированы с помощью специального программного обеспечения. Их также называют учебными [6], корпусами учебных текстов [7], корпусами ошибок [8] или корпусами студенческих текстов [9]. В мире в настоящее время существует несколько сотен учебных корпусов на 26 языках, и их количество постоянно растет, что свидетельствует об эффективности их использования в различных целях. Например, об использовании их при тестировании и оценке знаний пишут С. Гетц [10, с. 311–326], О. И. Виноградова [11], М. Боулес [12] и др. Кроме того, такие коллекции ученических текстов позволяют проводить исследования определенных тенденций в развитии иноязычных навыков у носителей того или иного языка [13, с. 3–18; 14, с. 5–22; 15, с. 79–97].

В лингводидактике особую значимость имеют лонгитюдные учебные корпуса, позволяющие исследовать процессы освоения языков. Е. Д. Смутьская в своей статье «Лонгитюдные исследования в лингвистике: опыт и перспективы» [16], описывая историю создания этой разновидности корпусов, отмечает, что первые лонгитюдные учебные корпуса текстов сформированы совсем недавно: если не учитывать международную систему обмена данными по детской речи CHILDES (Children Language Database Exchange System), возникшую в 1980-х гг. как «проброобраз лонгитюдных

¹ Под словом «аутентичные» подразумевается, что эти данные не были подвержены лингвистической рефлексии, т. е. могут содержать разного рода ошибки, которые зачастую и становятся предметом исследований.

корпусов текстов» [16, с. 55], первым таким опытом можно назвать проект Лувенского университета (Бельгия) 2008 г. под названием Longitudinal Database of Learner English (LONGDALE). В России лонгитюдные учебные корпуса тоже появились только недавно – первое описание в научных статьях относится к 2012 г. Е. Д. Смутьская описывает в указанной статье только три лонгитюдных корпуса носителей русского языка, о которых ей удалось найти хоть какие-то упоминания: «Корпус русской предпоздросковой речи», «Ваня» и проект SPbEFL LC, и делает вывод о том, что «в настоящее время лонгитюдные корпусные исследования востребованы, хотя проводятся гораздо реже синхронных». Причина этого очевидна: лонгитюдные исследования гораздо более затратны в плане ресурсов и времени. При этом нет сомнений, что по сравнению с синхронными исследованиями многолетние сборы данных обладают рядом явных преимуществ. Во-первых, лонгитюдное исследование позволяет представить процесс овладения языком в динамике. Во-вторых, оно дает возможность исследовать как тенденции в овладении языком в целом, так и индивидуальный стиль, динамику развития каждого обучающегося. Всем этим объясняется ценность электронных баз данных, создаваемых для таких лонгитюдных исследований.

В Петрозаводском государственном университете (ПетрГУ) в настоящее время ведутся работы по созданию, разметке и развитию Петрозаводского аннотированного корпуса текстов (ПАКТ), немецкоязычная часть которого по состоянию на конец 2022 г. включала более 1 100 студенческих текстов, написанных студентами ПетрГУ в период с 2018 г. по настоящее время. Корпус постоянно пополняется с целью получения обширной базы данных, пригодной для лонгитюдных исследований.

Целью данной статьи является описание возможностей, которые открывает ПАКТ для лонгитюдных исследований иноязычного образования у носителей русского языка. Данные возможности определяются в первую очередь разметкой (аннотациями), которая может отличаться от корпуса к корпусу.

ПАКТ имеет несколько видов аннотаций – автоматических и заполняемых вручную. К большей части текстов была применена автоматическая частеречная разметка, с помощью которой каждый токен (словоупотребление) в корпусе получил частеречную характеристику. Вручную размечаются лингвистические ошибки, для чего была разработана специальная иерархическая классификация, состоящая из 91 элемента. Все лингвистические ошибки распределяются по семи основным классам: грамматика, лексика, дискурс, пропуски, лишние элементы, орфография и пунктуация. Для каждой ошибки при разметке обязательно указываются ее тип, степень грубости и исправление. Факультативно, в случае отсутствия сомнений у аннотатора, размечается одна из возможных причин ошибки: интерференция с родным языком, интерференция с английским языком (первым иностранным языком студентов) или опечатка. Кроме того, каждый текст, прошедший аннотирование ошибок, получает оценку, учитывающую не только количество и степень грубости ошибок, но и содержательный и композиционный аспекты.

Очень важной для лонгитюдных исследований является метаразметка, сопровождающая каждый текст. В нее входят следующие пункты, заполняемые отчасти автоматически, отчасти вручную студентом, написавшим текст: автор; номер академической группы; на каком курсе написан текст; дата написания текста; самооценивание текста; понравилось ли задание; эмоциональное состояние в момент написания текста; тип текста (письмо, эссе, пересказ и т. п.); тип выполнения работы – на занятии/дома.

Такое разнообразие данных, собираемых на протяжении нескольких лет обучения в вузе, позволяет проводить лонгитюдные исследования динамики или взаимозависимостей по любому из указанных аспектов и их сочетаниям. В базе данных можно проследить, как меняется статистика по каждому из пунктов метаразметки (самооценивание, эмоциональное состояние, интерференция с родным или с английским языком и т. д.) с течением лет или в зависимости от разных параметров: пола, даты выполнения, темы и т. д. При этом можно отслеживать эти данные как для корпуса в целом, так и по курсам от 1 до 5 или для отдельно взятого студента.

В данной статье приведем пример одного из таких лонгитюдных исследований. Речь пойдет о статистике ошибок и ее динамике на разных курсах обучения немецкому языку. Важно отметить, что все обучающиеся, чьи работы собраны в корпусе ПАКТ, начинали изучать немецкий язык с нуля (в качестве основного иностранного языка в школе немецкий язык изучал только один студент).

Материал и методы

Благодаря структуре корпуса, где каждая работа отнесена к тому или иному курсу обучения, а также с помощью выполненной вручную разметки лингвистических ошибок по указанной выше классификации были получены данные по каждому курсу и году обучения с перечислением встретившихся типов ошибок и указанием их количества. На основе этих данных была составлена общая статистика ошибок сначала по 91 категории классификации.

Эта общая статистика показывает, что какие-то ошибки встречаются очень часто, а какие-то очень редко. Например, наиболее частыми являются ошибки на выбор лексики, орфографию, порядок слов, а наименее частыми являются, например, ошибки в числительных, в сравнительных конструкциях и в употреблении конъюнктива. Кроме того, во многих случаях в целом высокое или низкое число ошибок просматривается на протяжении всех пяти курсов, т. е. если лексических ошибок много на первом курсе, то их относительно много на протяжении всех курсов, подобно тому, как относительно низкое количество ошибок на использование пассивных конструкций остается незначительным также в течение всех пяти лет изучения языка. Однако при более внимательном взгляде становится очевидной динамика в каждом типе ошибок со своими пиковыми и провальными значениями. Например, несмотря на то что на первых четырех курсах ошибок с тегом «Стиль» почти нет, на последнем курсе этот показатель резко уходит в пиковые значения.

В ходе исследования мы обратились к десяти наиболее частым ошибкам первокурсников и проследили динамику этих ошибок в лонгитюдном срезе.

Результаты и обсуждение

Лексические ошибки

Лексические ошибки являются наиболее частым типом ошибок, при этом, как показывает статистика, именно эти ошибки чаще других оказываются критическими для понимания смысла. Если посмотреть на динамику количества таких ошибок на протяжении пяти лет обучения, то можно наблюдать, что этот показатель не только не становится меньше, но, начиная со второго курса, последовательно растет (рис. 1).

Такая динамика, возможно, связана с тем обстоятельством, что на старте изучения языка тексты, которые составляют студенты, базируются на определенном образце и конкретной лексике, которую они изучают. Лишь немногие авторы выходят за рамки шаблонов и составляют свои креативные сочинения, прибегая к помощи словарей и онлайн-переводчиков, что зачастую приводит к неверному выбору лексем. Чем старше становятся студенты, тем свободнее становятся их тексты, тем более сложные мысли они пытаются выразить, прибегая к более сложным синтаксическим конструкциям, требующим более разнообразной лексики. Опираясь на правила родного русского языка и не учитывая коллокационную сочетаемость, авторы обращаются за лексикой к современным онлайн-словарям, которыми, как видно по статистике ошибок, не умеют правильно пользоваться.

Таким образом, можно сделать вывод о необходимости специально обучать студентов работать с современными инструментами, прежде всего корпусными, позволяющими правильно подбирать лексемы с учетом коллокационной сочетаемости и нормами иностранного, а не своего родного языка.

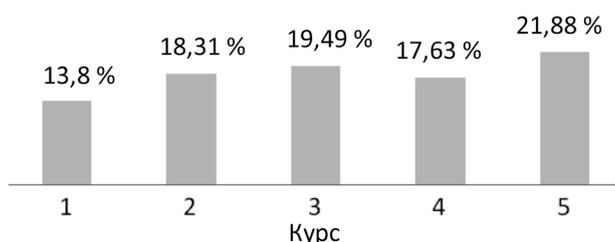


Рис. 1. Динамика лексических ошибок с первого по пятый курс

Неопределенный артикль

Следующим частотным типом ошибок в работах первокурсников являются ошибки на использование неопределенного артикля. Этот тег ставится в корпусе в том случае, если вместо нужного

неопределенного артикля стоит определенный или нулевой. В подавляющем большинстве случаев такие ошибки не влияют на понимание и вообще зачастую весьма спорны, так как разговорная форма теперь часто пренебрегает правилами употребления артиклей. Тем не менее в корпусе данные ошибки исправляются, хотя и имеют разметку как не влияющие на понимание.

Динамика ошибок на использование неопределенного артикля непоследовательная, что требует дополнительного анализа (рис. 2). Большое количество случаев неправильного употребления неопределенного артикля на начальном этапе обучения, вероятнее всего, связано с отсутствием такой грамматической категории в русском языке и существенными различиями в правилах употребления неопределенного артикля в немецком и английском, который студенты изучали в школе и который также может проявлять негативную интерференцию.

Резкое снижение показателя такого рода ошибок на втором курсе и последовательное снижение до незначительных 0,63 % на выпускном курсе можно было бы считать логичным результатом обучения, однако этому препятствует неожиданный скачок значения на четвертом курсе. Это увеличение, возможно, связано с тем, что в это время обучающиеся используют значительно больше описательных конструкций с большим количеством прилагательных в качестве атрибутов, перед которыми часто должен стоять неопределенный артикль, так как выраженный прилагательным атрибут подчеркивает характеристику объекта. Однако это пока лишь гипотеза, требующая подробного аналитического исследования.

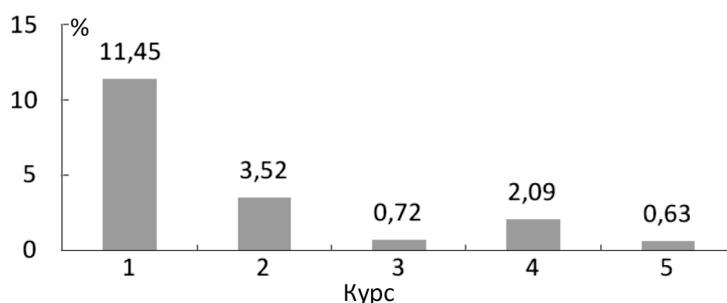


Рис. 2. Динамика ошибок на неопределенный артикль с первого по пятый курс

Орфографические ошибки

Почти так же много у начинающих изучать немецкий язык студентов орфографических ошибок (рис. 3).

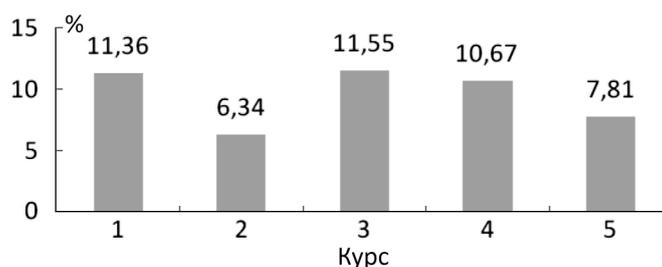


Рис. 3. Динамика орфографических ошибок с первого по пятый курс

В отличие от ошибок на неопределенный артикль, орфографические сохраняют статус наиболее частых на протяжении всех лет обучения. Следует отметить, что в орфографических ошибках наблюдаются явные тенденции, так что можно выделить большие группы: написание существительных со строчной буквы, написание местоимения «я» с прописной буквы, англоязычные варианты написания слов, похожих по звучанию с немецкими (например, Music), неверные написания имен собственных и печатки. Первые три указанные группы, напрямую связанные с интерференцией с английским как первым иностранным языком наших студентов, уходят к последнему курсу, так что на пятом курсе почти все орфографические ошибки связаны с опечатками по невнимательности. Это подтверждает и статистика: 11,36 % на первом курсе становятся 7,81 % на последнем курсе.

Ошибки на склонение прилагательных

Одной из наиболее образцовых диаграмм является диаграмма, отражающая статистику ошибок студентов на склонение прилагательных (рис. 4). На диаграмме видно, что на первом курсе это одна из самых частотных ошибок, но к пятому курсу их число падает до уровня ниже 1 %. При этом резкие скачки вниз наблюдаются на первых двух курсах. И хотя на диаграмме виден некоторый рост числа ошибок на склонение прилагательных в текстах студентов четвертого курса, все же этот рост очень незначительный и компенсируется спадом на пятом курсе.

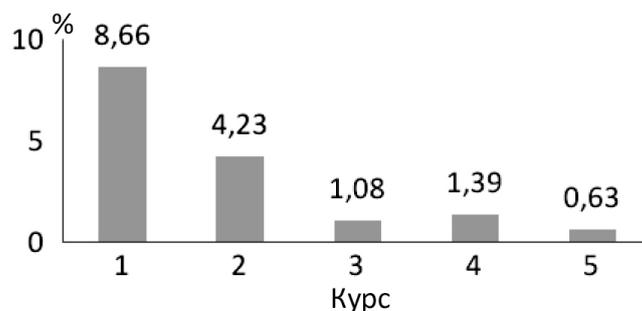


Рис. 4. Динамика ошибок на склонение прилагательных с первого по пятый курс

Пропуски

Замыкают пятерку самых частотных ошибок на первом курсе пропуски (рис. 5). Здесь наблюдается обратная динамика: к пятому курсу процент таких ошибок растет с 7,45 до 10,0 %, и они выходят у выпускников на второе место после лексических ошибок.

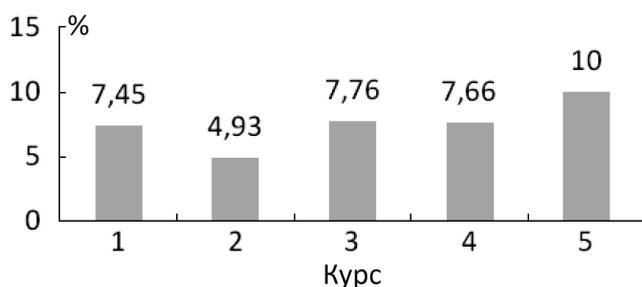


Рис. 5. Динамика ошибок на пропуски с первого по пятый курс

В студенческих текстах наблюдаются пропуски самых разных частей речи: частиц, возвратных местоимений, личных местоимений, союзов и полнозначных слов (рис. 6). Решение вопроса, с чем именно связана такая динамика, могло бы помочь найти способ, как изменить ситуацию, дать какой-то конкретный дидактический совет по профилактике роста подобных ошибок с увеличением словарного запаса и синтаксического инструментария.

3.123_4 , Dec. 11, 2020

Die Hauptfigur ist Lena , sie ist nicht wie in der Klasse und verspottet .

3.109_2 , Dec. 11, 2020

Aber ich denke , dieser Satz bedeutet , dass wenn man gesund , dann glücklich ist .

3.10_2

Und sie werden das Interesse am Lernen .

Рис. 6. Примеры предложений с ошибками на пропуск

Склонение имен существительных

На первом году обучения много ошибок в работах студентов связано со склонением имен существительных – 5,87 % от общего числа ошибок (рис. 7). На втором курсе этот показатель несколько растет, однако к пятому курсу последовательно падает до 0,6 %. Такая динамика вполне

логична, если учесть, что тегом ошибки «Склонение имен существительных» в корпусе отмечались также ошибки в склонении субстантивированных прилагательных, которые начинают активно использоваться студентами к концу первого курса.

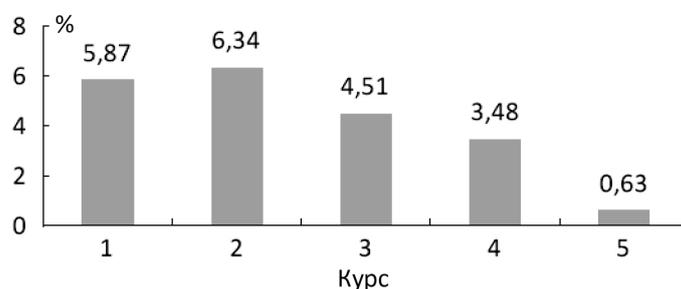


Рис. 7. Динамика ошибок на склонение существительных с первого по пятый курс

Пунктуационные ошибки

А вот ошибки на пунктуацию в продольном срезе пяти лет обучения показывают весьма странную динамику (рис. 8).

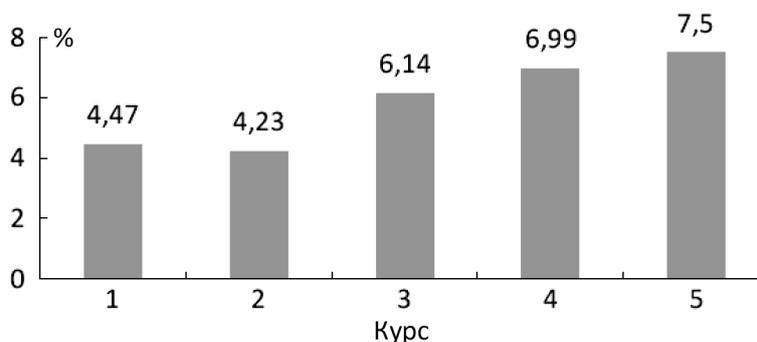


Рис. 8. Динамика пунктуационных ошибок с первого по пятый курс

Диаграмма иллюстрирует рост пунктуационных ошибок с 4,47 % на первом курсе до 7,5 % на пятом. С чем связан этот рост, еще только предстоит выяснить в ходе анализа пунктуационных ошибок отдельно на каждом курсе. Однако уже проведенное исследование [17] показало, что очень высокий показатель пунктуационных ошибок во всем корпусе в целом связан с интерференцией с русским языком, когда студенты выделяют запятыми слова *Zum Beispiel*, *Außerdem*, *auch*, *erstens*, *zweitens*. Возможно, рост пунктуационных ошибок связан с увеличением синтаксической сложности предложений на каждом последующем курсе.

Ошибки на образование множественного числа

Одна из самых непоследовательных динамик представлена в диаграмме ошибок на множественное число (рис. 9). Как и в случае с ошибками на склонение существительных, здесь наблюдается небольшой рост на втором курсе.

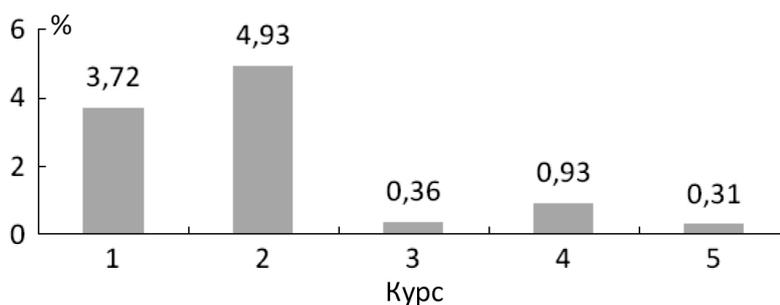


Рис. 9. Динамика ошибок на множественное число с первого по пятый курс

Возможно, это связано с тем, что в эту выборку попадают и сочинения, написанные в начале второго курса, т. е. когда знания соответствуют уровню первого курса, точнее, остаткам этих знаний после лета. Начиная с третьего курса мы видим резкое снижение до несущественных значений менее процента. Это можно расценивать как показатель того, что данная тема усваивается обучающимися хорошо.

Ошибки на род имени существительного

Не так хорошо даже к выпускному курсу студенты осваивают тему «Род имени существительного» (рис. 10). Только на третьем курсе показатель ошибок падает до уровня ниже процента, но затем несколько вырастает, при этом рост, хоть и незначительный, но все же рост, наблюдается и на четвертом, и на пятом курсах. В каких именно словах наблюдаются ошибки на разных курсах, можно ли проследить какие-то тенденции и дать конкретные методические рекомендации – темы будущих исследований.

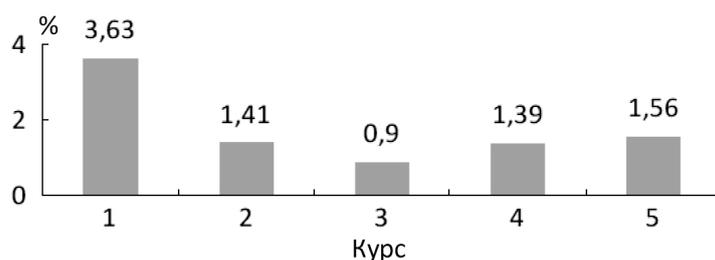


Рис. 10. Динамика ошибок на род имени существительного с первого по пятый курс

Ошибки на обратный порядок слов

Замыкают десятку самых распространенных на первом курсе ошибок нарушения в порядке слов. Следует указать, что для ошибок, связанных с порядком слов, в классификации корпуса имеются восемь тегов: помимо общих тегов ошибок на место глагола в предложении и ошибок на место других частей речи, туда входят более узко обозначенные ошибки на прямой порядок слов, на обратный порядок слов, на порядок слов в придаточных предложениях, на ошибки в рамочных конструкциях, на неверное тема-рематическое членение и на неверную постановку отрицания. Согласно статистике, на первом курсе ошибки на обратный порядок слов входят в десятку самых частотных ошибок (хотя процент все же уже невысокий – 3,3 %), и затем эта цифра последовательно снижается вплоть до пятого курса (рис. 11). С чем связан подъем процента ошибок на последнем курсе, предстоит еще проанализировать, но в любом случае этот процент ниже уровня первого курса.

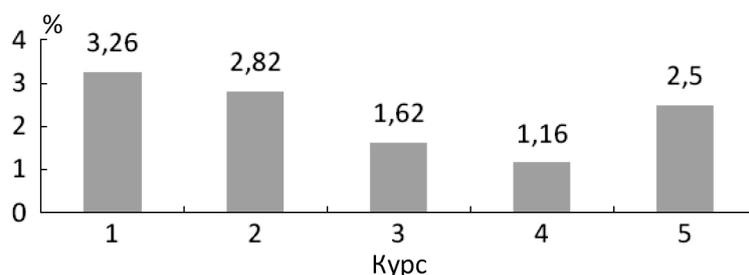


Рис. 11. Динамика ошибок на обратный порядок слов с первого по пятый курс

Выводы

Сравнение наиболее частотных ошибок на первом и пятом курсах (таблица) показывает, что к темам, которые на протяжении всех лет обучения вызывают наибольшие проблемы у студентов, относятся выбор лексемы, орфография, пропуски, пунктуация и обратный порядок слов. Уходят к концу обучения в вузе проблемы с неопределенным артиклем, склонением прилагательных и существительных, образованием множественного числа и родом существительных. Однако проявляются другие проблемы, такие как лишние элементы, логика, порядок слов в придаточных предложениях и стилистические ошибки.

Сравнение наиболее частотных ошибок на первом и пятом курсах

Наиболее частотные ошибки на первом курсе	Наиболее частотные ошибки на пятом курсе
1. Выбор лексемы	1. Выбор лексемы
2. Неопределенный артикль	2. Пропуски
3. Орфография	3. Орфография
4. Склонение прилагательных	4. Пунктуация
5. Пропуски	5. Лишние элементы
6. Склонение существительных	6. Логика
7. Пунктуация	7. Порядок слов в сложноподчиненном предложении
8. Множественное число	8. Личные местоимения
9. Род имени существительного	9. Стиль
10. Обратный порядок слов	10. Обратный порядок слов

В продолжение данного исследования важно посмотреть на лонгитюдные показатели для разных групп с целью анализа, являются ли показатели в целом схожими, что отражает общий процесс обучения немецкому языку русскоязычными студентами, или все же каждая группа уникальна и динамика освоения тех или иных аспектов немецкого языка в них протекает по-разному. В любом случае можно говорить о том, что ученический корпус служит базой данных, на которых можно проводить серьезные статистические лингводидактические исследования, способные сделать обучение немецкому языку более эффективным.

СПИСОК ИСТОЧНИКОВ

1. Трудности и перспективы цифровой трансформации образования / под. ред. А. Ю. Уварова, И. Д. Фрумина. М.: НИУ ВШЭ, 2019. 344 с. DOI: 10.17323/978-5-7598-1990-5
2. Павлова О. Ю. Использование языковых корпусов в обучении иностранному языку // Язык и культура. 2021. № 54. С. 283–298. DOI: 10.17223/19996195/54/16
3. Большие данные в образовании: доказательное развитие образования / под ред. О. А. Фиофановой. М.: Дело. 2021. 342 с.
4. Modeling Language Learning Using Specialized Elo Ratings. Innovative Use of NLP for Building Educational Applications / J. Hou, M. Koppatz, J. M. Hoya Quecedo, N. Stoyanova, M. Kopotev, R. Yangarber; eds. N. Yannakoudakis, E. Kochmar, C. Leacock, N. Madhani, I. Pilán, T. Zesch. Stroudsburg, PA: Association for Computational Linguistics, 2019. P. 494–506. <http://dx.doi.org/10.18653/v1/W19-4451>
5. Granger S. The International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research // TESOL Quarterly. 2003. Vol. 37, № 3. P. 538–546.
6. Камшилова О. Н. Учебный корпус текстов: потенциал, состав, структура. СПб.: Книжный дом, 2012. 56 с.
7. Ахапкина Я. Э. Эрратологическая разметка корпуса русских учебных текстов: тактические решения // Труды Института русского языка им. В. В. Виноградова. 2019. № 4 (22). С. 9–21.
8. 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. URL: <https://cyberleninka.ru/article/n/korpusy-oshibok-tselevaya-auditoriya-vozmozhnaya-arhitektura-korpusa> (дата обращения: 21.01.2023).
9. Котюрова И. А., Щеголева Л. В. Корпус студенческих текстов на немецком языке как источник данных для образования и науки // Вопросы образования. 2022. № 4. С. 322–349. URL: <https://cyberleninka.ru/article/n/korpus-studencheskih-tekstov-na-nemetskom-yazyke-kak-istochnik-dannyh-dlya-obrazovaniya-i-nauki> (дата обращения: 21.01.2023).
10. Götz S. Learner Corpora to Inform Testing and Assessment // The Routledge Handbook of Corpora and English Language Teaching and Learning. Routledge, 2022. P. 311–326.
11. Vinogradova O., Login N. The Design of Tests with Multiple Choice Questions Automatically Generated from Essays in a Learner Corpus // Higher School of Economics Research Paper No. WP BRP. 2017. Vol. 60. P. 16.
12. Bowles M. A. Using instructor judgment, learner corpora, and DIF to develop a placement test for Spanish L2 and heritage learners // Language Testing. 2022. Vol. 39, № 3. P. 355–376. <https://doi.org/10.1177/02655322221076033>

13. Granger S. The computer learner corpus: a versatile new source of data for SLA research // *Learner English on computer*. Routledge, 2014. P. 3–18.
14. Kwon H. English learner corpora and research in Korea // *Corpora*. 2022. Vol. 17, № Supplement. P. 5–22.
15. Liu K., Oiwan Cheung J., Zhao N. Learner corpus research in Hong Kong: past, present and future // *Corpora*. 2022. Vol. 17, № Supplement. P. 79–97.
16. Смутьская Е. Д. Лонгитюдные исследования в лингвистике: опыт и перспективы // *Известия РГПУ им. А. И. Герцена*. 2016. № 182. С. 53–58.
17. Сибирякова Н. Б. Интерференция с родным (русским) языком при пунктуационных ошибках в текстах студентов на немецком языке // *Языковые контакты в поликультурном мире*. Курск, 2022. С. 87–94.

References

1. Uvarov A. Yu., Frumin I. D. (eds) *Trudnosti i perspektivy tsifrovoy transformatsii obrazovaniya* [Challenges and prospects for digital transformation of education]. Moscow, NIU VSHE Publ., 2019. 344 p. (in Russian). DOI: <https://doi.org/10.17323/978-5-7598-1990-5>
2. Pavlova O. Yu. Ispol'zovaniye yazykovykh korpusov v obuchenii inostrannomu yazyku [Linguistic corpora in foreign language teaching]. *Yazyk i kul'tura – Language und Culture*, 2021, no. 54, pp. 283–298 (in Russian). DOI: <https://doi.org/10.17223/19996195/54/16>
3. Fiofanova O. A. (ed.) *Bol'shiye dannyye v obrazovanii: dokazatel'noye razvitiye obrazovaniya* [Big Data in Education: Evidence-based education development]. Moscow, Delo Publ., 2021. 342 p. (in Russian).
4. Hou J., Koppatz M., Hoya Quecedo J. M., Stoyanova N., Kopotev M., Yangarber R. Modeling Language Learning Using Specialized Elo Ratings. *Innovative Use of NLP for Building Educational Applications. Proceedings of the 14th Workshop*. Stroudsburg, PA: Association for Computational Linguistics, 2019. P. 494–506. DOI: <http://dx.doi.org/10.18653/v1/W19-4451>
5. Granger S. The International Corpus of Learner English: a new resource for foreign language learning and teaching and second language acquisition research. *TESOL Quarterly*, 2003, 37 (3), pp. 538–546.
6. Kamshilova O. N. *Uchebnyy korpus tekstov: potentsial, sostav, struktura* [Text corpora: potential, composition, structure]. Saint Petersburg, Knizhnyy dom Publ., 2012. 56 p. (in Russian).
7. Akhapkina Ya. E. Erratologicheskaya razmetka korpusa russkikh uchebnykh tekstov: takticheskiye resheniya [Erratorical marking of the textbook corpora: tactical solutions]. *Trudy Instituta russkogo yazyka im. V. V. Vinogradova – Proceedings of the V. V. Vinogradov Institute of Russian Language*, 2019, no. 4 (22), pp. 9–21 (in Russian).
8. Grudeva E. V., Buchilova I. A., Volkova N. A. Korpusy oshibok: tselevaya auditoriya, vozmozhnaya arkhitektura korpusa [Corpora of errors: target audience, a possible architecture of the corpus]. *Vestnik Cherepovetskogo gosudarstvennogo universiteta – Cherepovets State University Bulletin*, 2018, vol. 5 (86), pp. 63–72 (in Russian). URL: <https://cyberleninka.ru/article/n/korpusy-oshibok-tselevaya-auditoriya-vozmozhnaya-arhitektura-korpusa> (accessed 21 January 2023).
9. Kotyurova I. A., Shchegoleva L. V. Korpus studencheskikh tekstov na nemetskom yazyke kak istochnik dannykh dlya obrazovaniya i nauki [Learner corpus in German as a data source for education and science]. *Voprosy obrazovaniya – Educational Studies Moscow*, 2022, no. 4, pp. 322–349 (in Russian). URL: <https://cyberleninka.ru/article/n/korpus-studencheskikh-tekstov-na-nemetskom-yazyke-kak-istochnik-dannyh-dlya-obrazovaniya-i-nauki> (accessed 21 January 2023).
10. Götz S. Learner Corpora to Inform Testing and Assessment. *The Routledge Handbook of Corpora and English Language Teaching and Learning*, Routledge, 2022. Pp. 311–326.
11. Vinogradova O., Login N. The Design of Tests with Multiple Choice Questions Automatically Generated from Essays in a Learner Corpus. *Higher School of Economics Research Paper*, No. WP BRP, 2017, vol. 60, pp. 16.
12. Bowles M. A. Using instructor judgment, learner corpora, and DIF to develop a placement test for Spanish L2 and heritage learners. *Language Testing*, 2022, no. 39 (3), pp. 355–376. DOI: <https://doi.org/10.1177/02655322221076033>
13. Granger S. The computer learner corpus: a versatile new source of data for SLA research. *Learner English on computer*, Routledge, 2014. P. 3–18.
14. Kwon H. English learner corpora and research in Korea. *Corpora*, 2022, vol. 17, no. Supplement, pp. 5–22.
15. Liu K., Oiwan Cheung J., Zhao N. Learner corpus research in Hong Kong: past, present and future. *Corpora*, 2022, vol. 17, no. Supplement, pp. 79–97.

16. Smul'skaya E. D. Longitudynnye issledovaniya v lingvistike: opyt i perspektivy [Longitudinal Studies in Linguistics: Experience and Prospects]. *Izvestiya RGPU im. A. I. Gertsena – Herzen University Journal of Humanities and Sciences*, 2016, no. 182, pp. 53–58 (in Russian).
17. Sibiryakova N. B. Interferentsiya s rodnym (russkim) yazykom pri punktuatsionnykh oshibkakh v tekstakh studentov na nemetskom yazyke [Interference with the mother language (Russian) in punctuation errors in student texts in German]. *Yazykovyye kontakty v polikul'turnom mire* [Language contacts in a multicultural world]. Kursk, 2022. Pp. 87–94 (in Russian).

Информация об авторах

Котюрова И. А., кандидат филологических наук, доцент, Петрозаводский государственный университет (пр. Ленина, 33, Петрозаводск, Россия, 185001).

Сафонов Г. Р., студент, Петрозаводский государственный университет (пр. Ленина, 33, Петрозаводск, Россия, 185001).

Information about the authors

Kotyurova I. A., Candidate of Philological Sciences, Associate Professor, Head of the Department of German and French Languages, Petrozavodsk State University (ul. Lenina, 33, Petrozavodsk, Russian Federation, 185910).

Safonov G. R., undergraduate student, Petrozavodsk State University (ul. Lenina, 33, Petrozavodsk, Russian Federation, 185910).

Статья поступила в редакцию 06.03.2023; принята к публикации 22.12.2023

The article was submitted 06.03.2023; accepted for publication 22.12.2023