

<https://www.doi.org/10.33910/1992-6464-2022-203-247-254>

Хоу Линьсюе

КЛАССИФИКАЦИЯ ОШИБОК ДЛЯ СОЗДАНИЯ РАЗМЕТКИ В УЧЕБНОМ КОРПУСЕ РУССКИХ ТЕКСТОВ

В статье рассматривается актуальная проблема классификации ошибок, которая является основой для разметки учебного корпуса текстов. Определяется типология ошибок, возникающих при овладении иностранным языком, анализируются разные системы классификации ошибок в русском языке. Основное внимание уделяется различным типам письменных ошибок, допускаемых китайскими студентами при овладении русским языком, и их классификации для разметки в учебном корпусе.

Ключевые слова: учебный корпус, понятие ошибки, классификация ошибок, РКИ, китайские студенты

Hou Linxue

CLASSIFICATION OF ERRORS IN RUSSIAN TEXTS FOR BUILDING TAGGED LEARNER CORPORA

The paper discusses approaches to the classification of errors as the basis for the development of a tagged learner corpora. The paper provides a definition of error in foreign language learning. It also analyses different classifications of errors in the Russian language. A special focus is given to different types of writing errors made by Chinese students of Russian and classification of these errors for the development of tagged learner corpora.

Keywords: learner corpora, definition of error, error classification, Russian as foreign language, Chinese students

По мере развития и совершенствования информационных технологий, применяемых в филологических науках, особое место такого относительно нового направления как корпусная лингвистика, занимает учебный корпус, который доказал возможность своего использования при обучении родным и иностранным языкам [5, с. 56]. Используемый как база для одного из новых методов обучения, учебный корпус сосредоточивается на мониторинге и анализе ошибок учащихся в процессе овладения иностранным языком. Можно сказать, что учебный корпус строится на основе данных об ошибках. С одной стороны, точность и адекватность классификации и идентификации ошибок определяют эффективность использования учебного корпуса, с другой стороны, они

представляют особую трудность для создания корпуса такого типа. Поэтому создание учебного корпуса текстов китайских студентов на русском языке требует предварительного рассмотрения общего подхода к определению того, что понимается под этим термином, и классификации ошибок, необходимой для их разметки в корпусе.

Для того, чтобы классифицировать ошибки, необходимо определить, что понимается под этим термином. Ошибка появляется в любой сфере человеческой деятельности, но в разных научно-исследовательских сферах и предметных областях она определяется по-разному. В данной статье рассматриваются ошибки, возникающие при овладении иностранным языком.

Широкое толкование понятия «*ошибка*» определяется в свете отклонения от стандарта и межъязыковой интерференции [9, с. 390]. С точки зрения отклонения от стандарта ошибка рассматривается как нарушения правильного использования языковых единиц и их форм, т. е. результат ошибочного действия [1, с. 182]. При этом отклонением от стандартизованного использования является некорректный выбор единицы из ряда одноуровневых единиц или члена одной парадигмы для данных условий функционирования [6, с. 28]. Отсюда следует, что неверное действие учащегося приводит к ошибке, а ошибка является проявлением речевого и мыслительного действий студентов.

Ошибка может возникать под влиянием межъязыковой интерференции, под которой понимается расхождение с нормой в системе изучаемого языка под влиянием языка родного [2, с. 81]. Явление межъязыковой интерференции является одной из основных причин ошибок. В этом случае ошибка определяется как явление, связанное с отрицательным влиянием знаний родного языка на знание изучаемого языка. Из-за трудности усвоения абсолютно новых, ранее в родном языке не встречавшихся, явлений изучаемого языка обучающийся допускает множество ошибок при овладении им.

Анализируя ошибки учащихся, можно выявить трудности, возникающие в процессе овладения иностранным языком и учесть их при определении методической направленности для преподавателя и улучшения качества преподавания. Однако при таком подходе нет точного критерия для оценки отклонения учащегося от стандарта и влияния именно интерференции, в большинстве случаев установление ошибки зависит от личного опыта, поэтому является субъективным.

Узкое толкование понятия «*ошибка*» связано с нарушением норм [9, с. 390]. При этом принято определять ошибку с позиции преподавателя и с позиции носителя языка. С точки зрения преподавателя ошибкой является нарушение нормы, которая не только положена в основу учебников по грамматике, но и существует в сознании преподавателя, то есть то, что

он считает неправильным, оценивается им как ошибка [8, с. 65]. Однако и здесь имеются свои недостатки: идентификация ошибки опирается на мнение преподавателя, но трудно объяснить, кто контролирует норму, на которую опирается преподаватель, и как эта норма возникла.

По сравнению с позицией преподавателя определение ошибки с позиции носителя языка более объективно и легче применяется в практических целях. В отношении классификации ошибок, допускаемых учащимися, целесообразна позиция С. Джеймса: «ошибка — нарушение норм речи носителя языка» [10, с. 67]. На этой основе можно определить ошибку по трем параметрам: грамматичность (правильность), приемлемость (понятность) и применимость:

- 1) грамматичность (правильность), ошибка — нарушение требований грамматики, например, орфографическая ошибка, лексическая ошибка и т. д.;
- 2) приемлемость (понятность): ошибка — это то, чего не понимает и принимает носитель языка, например, семантическая ошибка;
- 3) применимость: ошибка — неправильное использование стиля речи, например, использование в научном тексте разговорных выражений.

Основной целью работы является корпусное исследование ошибок в русском языке, которые допускают китайские студенты, и создание правил их разметки в корпусе. На основе рассмотренного выше определения «*ошибка*» понимается как нарушение грамматичности, правильности и приемлемости речи с точки зрения носителя русского языка. Иначе говоря, ошибка в речи китайских студентов, противоречащая правилам и нормам русского языка, воспринимается как неправильность.

Вопрос о том, соответствуют ли используемые слова требованиям стиля речи, выходит за рамки данного исследования, поскольку определение таких ошибок более субъективно. Следует отметить, что идентификация ошибок в основном зависит от грамматичности и правильности, остальные критерии играют

решающую роль только в определенных случаях.

В современной методике обучения русско-му языку как иностранному принято выделять несколько видов ошибок, допускаемых студентами. Наиболее распространенная классификация ошибок в речи иностранных студентов основана на следующих аспектах языка: словообразовательная ошибка (неоправданное сочинительство); морфологическая ошибка (ненормативное образование слов и употребление частей речи); синтаксическая ошибка (неверное построение предложений); лексическая ошибка (нарушение лексической сочетаемости) и стилистическая ошибка (нарушение единства стиля языка) [3, с. 81].

Опираясь на классификацию ошибок по аспектам языка будем отдельно рассматривать ошибки в использовании каждой части речи для того, чтобы выявить конкретные трудности китайских учащихся при использовании частей речи и их языковые характеристики при овладении русским языком. Дело в том, что такая классификация в основном связана с различием между китайским и русским языками и механизмом разметки ошибки. С одной стороны, между китайским и русским языками существуют большие различия. Русский язык — синтетический, то есть в нем преобладают синтетические формы выражения грамматических значений [4, с. 131]. Грамматические значения слова обычно передаются с помощью необходимых форм (вид, род, спряжение, управление и т. д.). В связи с этим русский язык сильно отличается от китайского языка, передающего грамматические отношения через отдельные служебные слова, через фиксированный порядок слов и контекст. Из-за этих различий китайские учащиеся часто встречаются с трудностями и допускают больше ошибок в использовании частей речи русского языка. И эти ошибки заслуживают особого внимания при классификации и анализе ошибок, допускаемых китайскими учащимися, они должны быть особым образом размечены в корпусе.

С другой стороны, в настоящее время механизм разметки корпуса может быть автоматическим, ручным или автоматическим с ручной корректировкой [7, с. 48]. Среди этих методов наиболее часто используется является ручная разметка, которая и выбрана для разметки создаваемого корпуса текстов. Для ручной разметки следует учитывать следующее: нельзя сводить множество типов ошибок, выявленных в конкретном месте текста, к использованию одного идентификатора — тега. Если один тег характеризует много ошибок, трудно выделить характеристики распространения ошибок учащихся, это может повлиять на анализ конкретных трудностей в обучении. Кроме того, следует учитывать, что детальность классификации ошибок способствует ее информативности, но нужно принимать во внимание, что стремление к детальной классификации ошибок приводит к появлению слишком большого количества тегов. Это не только усложняет ручную разметку, но и не способствует ее практическому использованию.

В качестве материала предварительного исследования были использованы 164 сочинения 41 студента 1-го курса по специальности русский язык Второго Пекинского университета иностранных языков. Общий объем этого пилотного массива составляет 26,4 тыс. токенов — слогоморфем.

Учитывая все сказанное выше, а также проведенное предварительное исследование массива текстов, можно разделить ошибки китайских учащихся на 4 основных типа и 14 подкатегорий ошибок, также проиллюстрировать ошибки примерами для детального описания. При разметке все анализируемые типы ошибок в корпусе отмечены строчными латинскими буквами и знаком <...>, а сама ошибка выделяется жирным шрифтом, например: *Мы должны использовать **каждый** возможность <пр2> учиться.*

Перечень типов ошибок представлен в таблице 1.

Типы ошибок для разметки в учебном корпусе (предварительный вариант)

Тип ошибки	Подкатегория ошибки	Тег	Название ошибки	Пример
морфологическая ошибка	орфографическая ошибка	fm1	орфографическая ошибка за счет незнания орфограмм	много <fm1>; правильный <fm1>; претсдвляет <fm1>; английский <fm1>; учят <fm1>
		fm2	неверное словообразование	прийдти<fm2>; весвый<fm2>; супермаркет<fm2>
		fm3	неверное использование прописной буквы	Я поздравляю вас с днём<fm3> Победы!
	ошибка в использовании глаголов	vr1	неверное использование возвратной формы	Я уже два года не виделась <vr1> друзей.
		vr2	неверное использование форм несовершенного вида	Нельзя проводить <vr2> футбольный матч, потому что завтра будет дождь. Я хочу изучать <vr2> русский язык в течение двух лет.
		vr3	неверное использование форм совершенного вида	Давайте защитить <vr3> окружающую среду вместе. В университете я продолжаю изучить <vr3> русский язык.
		vr4	неверное управление глагола	Родители должны уделять большое внимание своих детей <vr4>.
		vr5	неверное использование форм деепричастия	Выйдя <vr5> из метро, пошёл дождь.
		vr6	неверный выбор типа причастия	Из всех языков, изучающих <vr6> в университете, студенты часто выбирают английский язык.
	ошибка в использовании имен существительных	nr1	неверное использование форм числа имен существительных	Родители стараются создать хорошее условие <nr1>, чтобы дети получили хорошее образование.
		nr2	отсутствие согласования по роду	Мы должны использовать каждый возможность <nr2> учиться.
		nr3	неверное образование форм склонения	Студенты из деревен <nr3> учатся лучше, чем студенты из городов.
		nr4	нарушение падежных форм	Обеспечить школам <nr4> достаточным количеством тепла.
	ошибка в использовании предлогов	pp1	неверное управление предлога	Мая мама в свободное время часто гуляет по магазинами <pp1>.
		pp2	использование неверного предлога	В <pp2> третьем курсе передо мной появилась куча новых вопросов: много новых слов в текстах, чтение статей.
	ошибка в использовании имен прилагательных	aj1	неверное использование полной формы	Из-за того, что мой русский язык слабый <aj1>, я решила стараться учиться.
aj2		неверное использование степеней сравнения	Русский язык более труднее <aj2> английского.	

Таблица 1. Продолжение

	ошибка в использовании местоимений	pr1	неверный выбор местоимения	Сейчас, без интернета мы совсем не будем знать, что случилось и в своей <pr1> стране, и в мире.
	ошибка в использовании наречий	ad1	неверное использование степеней сравнения	Он говорит по-русски лучше <ad1> в нашей группе.
	ошибка в использовании числительных	nu1	неверное использование форм числительных	В каждой группе работают двое <nu1> преподавателей.
лексическая ошибка	ошибка в использовании слова	wd1	употребление слова в несвойственном ему значении	Сейчас большинство мальчиков ценят девушку только по красивому лицу и стройной внешности <wd1>.
		wd2	тавтология и повтор	Следует различать различные <wd2> подходы к проблеме.
		wd3	смещение синонимов и паронимов	Я хочу вступить <wd3> в аспирантуру после окончания университета. Брови его удивительно <wd3> поднялись.
	неверное использование устойчивых словосочетаний	cc1	неверное использование устойчивых выражений	получение больших успехов <cc1>; знания возросли <cc1>; взять образование <cc1>; крепко понять <cc1>; высокое образование <cc1>; совсем ленивый <cc1>
		cc2	фразеологическая ошибка	Тебе нельзя сидеть сложив руки <cc2> и только ожидать помощи.
синтаксическая ошибка	нарушение структуры предложений	sn1	пропуск членов предложения	На метро нужно <sn1> 30 минут из торгового центра до нашего <fm1> университета. Они сидели ха компьютером целыми часами, как известно, <sn1> это плохое для здоровья.
		sn2	неверный порядок слов	Главное - подаренные <sn2> цветы с любовью.
	нарушение норм формального синтаксиса	sn3	нарушение согласования подлежащего и сказуемого	Она улыбнулась и сказал <sn3>, что ей больше нравится заниматься спортом.
		sn4	нарушение видо-временной соотнесенности однородных членов предложения или сказуемых в главном и придаточном предложениях	По мере того как развиваются наука и техника, рос <sn4> поток информации.
	пунктуационная ошибка	pt1	отсутствие знака препинания	На зимних каникулах я заметил <pt1> что сосед купил сыну велосипед.
		pt2	лишний знак препинания	В университете, <pt2> мы часто помогаем друг другу.
		pt3	неверное использование знака препинания	Китайская кухня, <pt3> это самая острая кухня в мире.

Таблица 1. Продолжение

коммуникативная ошибка	логическое нарушение	lc1	нарушение логики при перечислении	На уроке присутствовали директор, профессора, также Анна Петровна Иванова, Чжан Цин <lc1>.
		lc2	отсутствие формальных показателей связи между предложениями	Мир развивается. Страна развивается. Город развивается. Во многих областях <pr3>нужна вода...<lc1>

Проанализируем процентное соотношение разных типов ошибок в учебном корпусе (табл. 2).

На основе списка, представленного в таблице 2, можно сделать вывод, что самой распространенной ошибкой, допускаемой

Таблица 2

Процентное соотношение разных типов ошибок в учебном корпусе

Тип ошибки	Подкатегория ошибки	Процент (%)
морфологическая ошибка	орфографическая ошибка	5,7
	ошибка в использовании частей речи	29,9
лексическая ошибка	ошибка в использовании слова	22,8
	неверное использование устойчивых словосочетаний	11,4
синтаксическая ошибка	нарушение структуры предложений	9,1
	нарушение норм формального синтаксиса	6,6
	пунктуационная ошибка	9,8
коммуникативная ошибка	логическое нарушение	4,7

китайскими студентами, является морфологическая ошибка, которая в основном встречается в использовании частей речи. Список, ранжированный по проценту ошибок в использовании частей речи, представлен в таблице 3. Согласно данным таблицы 3, наиболее сложными словами для китайских студентов оказались глаголы.

Проведенный анализ частоты ошибок в учебном корпусе, показал, что для китайских студентов трудностью при овладении русским языком является использование частей речи, особенно использование глаголов.

Анализ пилотного варианта учебного корпуса показал, что для преподавателей ошибка в использовании частей речи также представляет методический центр тяжести. При этом они должны делать упор на употребление частей речи, составлять разные виды упражнений с целью помощи студентам преодолеть эту трудность в процессе овладения русским языком.

Опираясь на приведенную выше классификацию ошибок, можно с помощью учебного корпуса точно анализировать трудности и языковые характеристики китайских

Процентное соотношение ошибок в употреблении частей речи

Название ошибки	Процент (%)
ошибка в использовании глаголов	40,2
ошибка в использовании имен существительных	29,8
ошибка в использовании предлогов	14,2
ошибка в использовании имен прилагательных	7,4
ошибка в использовании местоимений	3,8
ошибка в использовании числительных	3,2
ошибка в использовании наречий	1,4

студентов при овладении русским языком. Кроме того, используя результат анализа ошибок на занятиях, можно не только оптимизировать процесс преподавания и улуч-

шить качество овладения языком, но также помочь студентам преодолеть трудности изучения языка и повысить эффективность их обучения.

СПИСОК ЛИТЕРАТУРЫ

1. Азимов Э. Г., Щукин А. Н. Новый словарь методических терминов и понятий (теория и практика обучения языкам). М.: ИКАР, 2009. 448 с.
2. Гасконь Е. А. Некоторые ошибки китайских студентов, связанные с влиянием ранее освоенных языков на изучение нового (русского как иностранного) // Филологические науки. Вопросы теории и практики. 2016. № 10–3 (64). С. 80–84.
3. Жендаренко Е. В. Типичные ошибки, выявленные при тестировании по русскому языку как иностранному // В мире науки и искусства: вопросы филологии, искусствоведения и культурологии. 2013. № 29. С. 80–84.
4. Злобина Ю. И., Сорокина М. О. Некоторые типичные ошибки в письменных работах китайских студентов: классификация и анализ // Мир науки, культуры, образования. 2018. № 4 (71). С. 130–132.
5. Камшилова О. Н. Корпусные технологии в обучении языкам // Новые технологии в образовательном пространстве родного и иностранного языка. 2013. № 1. С. 56–62.
6. Кондрашова Н. В. Прогнозирование и исправление студенческих ошибок при обучении иностранным языкам // Научный диалог. 2015. № 7 (43). С. 27–47.
7. Митрофанова О. А., Захаров В. П., Азарова И. В., Попов А. М., Хохлова М. В. Моделирование в корпусной лингвистике: специализированные корпуса русского языка. СПб.: Изд-во СПбГУ, 2019. 208 с.
8. Морозова С. А., Лопинова М. Н., Миронова Н. В., Кондратьева О. И. Понятие «ошибка», причины возникновения и классификация ошибок в методике преподавания иностранного языка // Известия Великолукской Государственной Сельскохозяйственной Академии. 2016. № 3. С. 63–71.
9. Щелокова А. А. Типологизация речевых ошибок на основе понятий «конструктивность/деструктивность» // Мир науки, культуры, образования. 2020. № 4 (83). С. 389–391. <https://doi.org/10.24411/1991-5497-2020-00802>
10. James C. Errors in language learning and use: Exploring error analysis. Beijing: Foreign Language Teaching and Research Press, 2001. 313 p.

REFERENCES

1. Azimo E. G., Shchukin A. N. Novyj slovar' metodicheskikh terminov i ponyatij (teoriya i praktika obucheniya yazykam). M.: IKAR, 2009. 448 s.

2. *Gaskon' E. A.* Nekotorye oshibki kitajskikh studentov, svyazannye s vliyaniem ranee osvoennykh yazykov na izuchenie novogo (russkogo kak inostrannogo) // *Filologicheskie nauki. Voprosy teorii i praktiki.* 2016. № 10–3 (64). S. 80–84.

3. *Zhendarenko E. V.* Tipichnye oshibki, vyyavlennye pri testirovanii po russkomu yazyku kak inostrannomu // *V mire nauki i iskusstva: voprosy filologii, iskusstvovedeniya i kul'turologii.* 2013. № 29. S. 80–84.

4. *Zlobina Yu. I., Sorokina M. O.* Nekotorye tipichnye oshibki v pis'mennykh rabotakh kitajskikh studentov: klassifikatsiya i analiz // *Mir nauki, kul'tury, obrazovaniya.* 2018. № 4 (71). S. 130–132.

5. *Kamshilova O. N.* Korpusnye tekhnologii v obuchenii yazykam // *Novye tekhnologii v obrazovatel'nom prostranstve rodnogo i inostrannogo yazyka.* 2013. № 1. S. 56–62.

6. *Kondrashova N. V.* Prognozirovaniye i ispravleniye studencheskikh oshibok pri obuchenii inostrannym yazykam // *Nauchnyj dialog.* 2015. № 7 (43). S. 27–47.

7. *Mitrofanova O. A., Zakharov V. P., Azarova I. V., Popov A. M., Khokhlova M. V.* Modelirovaniye v korpusnoj lingvistike: spetsializirovannyye korpusy russkogo yazyka. SPb.: Izd-vo SPbGU, 2019. 208 s.

8. *Morozova S. A., Lopinova M. N., Mironova N. V., Kondrat'eva O. I.* Ponyatie "oshibka", prichiny vozniknoveniya i klassifikatsiya oshibok v metodike prepodavaniya inostrannogo yazyka // *Izvestiya Velikolukskoj Gosudarstvennoj Sel'skokhozyajstvennoj Akademii.* 2016. № 3. S. 63–71.

9. *Shchelokova A. A.* Tipologizatsiya rechevykh oshibok na osnove ponyatij "konstruktivnost'/destruktivnost'" // *Mir nauki, kul'tury, obrazovaniya.* 2020. № 4 (83). S. 389–391. <https://doi.org/10.24411/1991-5497-2020-00802>

10. *James C.* Errors in language learning and use: Exploring error analysis. Beijing: Foreign Language Teaching and Research Press, 2001. 313 p.