

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
имени М.В. ЛОМОНОСОВА

На правах рукописи

Морозов Даниил Александрович

Корпусное исследование степеней сравнения чешских прилагательных

Специальность: 5.9.6. Языки народов зарубежных стран (славянские языки)

АВТОРЕФЕРАТ

диссертации на соискание ученой степени

кандидата филологических наук

Москва – 2026

Диссертация подготовлена на кафедре славянской филологии филологического факультета Московского государственного университета имени М.В. Ломоносова.

Научный руководитель:

Изотов Андрей Иванович
доктор филологических наук, доцент

Официальные оппоненты:

Кульпина Валентина Григорьевна
доктор филологических наук, доцент,
Институт научной информации по
общественным наукам Российской
академии наук, ведущий научный
сотрудник отдела языкознания

Бекасова Елена Николаевна
доктор филологических наук, доцент,
Оренбургский государственный
педагогический университет,
филологический факультет, профессор
кафедры русского языка и методики
преподавания русского языка

Ващенко Дарья Юрьевна
кандидат филологических наук
Институт славяноведения Российской
академии наук, старший научный
сотрудник отдела славянского языкознания

Защита диссертации состоится «7» октября 2026 г. в 16 часов 30 минут на заседании диссертационного совета МГУ.059.1 Московского государственного университета имени М.В. Ломоносова по адресу: 119991 ГСП-1, Москва, Ленинские горы, МГУ имени М.В. Ломоносова, 1-й корпус гуманитарных факультетов, филологический факультет, ауд. 1060.

E-mail: sovets@philol.msu.ru

С диссертацией можно ознакомиться в отделе диссертаций научной библиотеки МГУ имени М.В. Ломоносова (Ломоносовский просп., д. 27) и на портале <https://dissovet.msu.ru/dissertation/3968>.

Автореферат диссертации разослан «___» _____ 2026 г.

Ученый секретарь
диссертационного совета,
доктор филологических наук

Е.А. Кузьмина

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Настоящее исследование посвящено корпусному анализу форм степеней сравнения чешских прилагательных и выявлению специфики моделей образования компаративов и суперлативов. В частности, научный интерес представляет изучение синтетичности как характеристики градационности.

Первостепенные функции любого языка – хранение и передача информации, и реализовываться они могут различными способами. Морфологическая типология разделяет языки на две большие группы: синтетические и аналитические. В синтетических языках грамматическое значение выражается в пределах слова, то есть словоформами – при помощи флексий, аффиксов, ударения, супплетивизма. Они противопоставлены аналитическим структурам, выражающим значения служебными словами.

Однако естественные языки не являются типологически однородными. Обе группы – синтетические и аналитические – взаимосвязаны и оказывают влияние друг на друга. Языки славянской группы традиционно относят к синтетическим, однако черты синтетизма и аналитизма выражены в большей или меньшей степени в каждом конкретном языке. Так, болгарский язык утратил падежную систему, но это вовсе не означает, что он не может выразить субъектно-объектных отношений. В болгарском языке они выражаются через сочетания существительных с предлогами, то есть аналитически: *отивам във влака на вкъщи* «я еду поездом домой». Наоборот, чешский и русский языки – синтетические, с развитой флективной системой: *jedu vlakem domů* «еду поездом домой». Если в болгарском языке грамматические формы степеней сравнения представлены синтетическими конструкциями: *по-малък* «меньше», *най-висок* «выше», то в русском или польском частотны аналитические сочетания *более интересный*, *bardziej ciekawy*. Какое же положение занимает в данном случае чешский язык, где встречаются оба типа конструкций?

На общеславянском фоне чешский язык выделяется архаичными и уникальными чертами, исходящими из истории его геополитического развития.

В. Матезиус считал чешский язык «самым архаичным членом славянского языкового мира»¹, а Карел Чапек называл чешскую речь «самой трудной из всех», «из самых богатых всеми оттенками и значениями»². По сравнению с другими языками славянской группы чешский наиболее синтетичен³. Если сравнивать модели выражения степеней сравнения в чешском языке и, например, в русском или польском, это на первый взгляд действительно кажется правдоподобным. Интерес представляет определение степени синтетичности, что практически можно реализовать проведением исследования частотности синтетических и аналитических форм на материале языковых корпусов. Таким образом можно получить статистические данные употребления сложных и простых форм, отследить контексты и проверить утверждение о синтетическом характере чешского языка и – в частности – о градационности качественного признака чешских прилагательных. Важным аспектом такого исследования является выявление возможности сосуществования синтетических и аналитических форм, что не считается возможным с точки зрения авторов пражской «Русской грамматики» и едва ли распространено, учитывая приведенные в «Грамматике современного чешского языка» статистические данные.

Итак, в практической части данного исследования предлагается рассмотреть особенности градационности качественного признака чешских прилагательных на примере синтетических и аналитических моделей формирования степеней сравнения, а именно – проверить гипотезу о синтетичности моделей градационности и подтвердить ее данными статистических расчетов.

Актуальность. Настоящее исследование может считаться актуальным, так как – с одной стороны – наблюдается глобальное повышение внимания

¹ *Mathesius V.* Čeština a obecný jazykozpyt: soubor statí. V Praze: Melantrich, 1947.

² *Čapek K.* Marsyas. Jak se co dělá. 2. vyd. Marsyas, 3. vyd. Jak se co dělá. Praha: Československý spisovatel, 1984. 384 s. Spisy, sv. 13

³ *Рылов С.А.* Проблема типологии современных славянских языков: социально-функциональный аспект // Лингвистические традиции и современность: сб. ст., посвящ. 90-летию проф. В.Н. Немченко / отв. ред. Л.В. Рацибурская. Нижний Новгород: ННГУ, 2018. С. 113–120.

современных лингвистов к корпусным исследованиям, переводящим изучение языкового материала на принципиально иной уровень, а с другой – мощное развитие компьютерных инструментов лингвистических исследований и автоматической обработки естественного языка. Если ранее для сбора достоверных сведений об относительной частотности тех или иных форм необходимо было вручную проанализировать все вхождения на некотором объеме литературы, причем выбранной и сбалансированной с учетом ее количества и стилей, то в наши дни объемы корпусов, возможность применить фильтры в реальном времени и автоматизировать процесс однотипных действий – позволяют значительно ускорить анализ. В условиях современного развития технологий электронно-вычислительной техники обращение к корпусу как основному источнику фактического материала даже не в разы, но в порядки повышает надежность полученных данных. Это в особенности востребовано при проверке в корпусе спорных гипотез. Например, на материале корпуса Synek было показано, что аналитические компаративные и суперлативные конструкции прилагательных в современном чешском языке могут образовываться далеко не только тогда, когда образование соответствующих синтетических форм недопустимо, как считали авторы пражской «Русской грамматики» 1979 г. В вопросе градационности качественного признака прилагательных корпус текстов чешского языка позволяет получить данные о действительном соотношении ее моделей и реальные контексты употребления тех или иных форм.

Целью исследования стало выявление особенностей градационности качественного признака чешских прилагательных.

Достижение заявленной цели предполагало решение следующих **конкретных задач**:

- 1) определить понятие градационности качественного признака прилагательных;
- 2) проследить развитие общеславянской парадигмы градационности в чешском языке;

- 3) установить характер синтетизма чешских прилагательных на примере степеней сравнения;
- 4) вычислить частотность компаративных и суперлативных форм чешских прилагательных на материале Чешского национального корпуса;
- 5) выявить возможность конкуренции синтетических и аналитических форм чешских прилагательных и построить таблицы конкуренции таких форм;
- 6) выделить характерные черты моделей выражения степеней сравнения чешских прилагательных на общеславянском фоне.

Выявление возможности конкуренции синтетических и аналитических форм, а также выделение характерных черт градационности качественного признака в формах степеней сравнения требовало решения вспомогательных задач. В частности, путем анализа материала языковых корпусов следовало решить данные задачи:

- 1) вычислить частотность употребления синтетических и аналитических форм превосходной степени;
- 2) вычислить частотность употребления синтетических и аналитических форм сравнительной степени;
- 3) выявить возможность сосуществования и конкуренции простых и сложных форм степеней сравнения;
- 4) построить таблицы конкуренции и вычислить относительную частоту употребления синтетических форм на основании полученных статистических данных.

Объектом исследования являются формы степеней сравнения прилагательных, а **предметом** – частотность употребления синтетических и аналитических моделей выражения степеней сравнения.

Материал. Частотность функционирования форм степеней сравнения вычислялась на материале сбалансированных репрезентативных корпусов чешских письменных текстов серии SYN в составе Чешского национального

корпуса, а именно: Synek, SYN2000, SYN2005, SYN2010, SYN2015 и SYN2020⁴. В совокупности данные корпуса охватывают период с 1990 по 2019 гг., а потому отражают как изменения за последние тридцать лет, так и устоявшиеся тенденции в современном чешском письменном дискурсе.

Выбор в качестве эмпирического материала данных корпусов серии SYN объясняется принципами отбора входящих в них текстов, что и делает данные корпуса репрезентативными и сбалансированными. Так как письменный текст не только отражает языковую ситуацию, но и формирует ее, влияя на идиолект читателя (а через совокупность идиолектов – и на язык нации), для отбора текстов создатели корпусов серии SYN опирались на результаты социологических исследований о читательских предпочтениях современников. Так, например, в SYN2000 основную часть языкового материала составляют публицистические тексты (60%), на втором месте – специальные тексты, справочники, энциклопедии и т.д. (25%), на третьем – беллетристика (15%). В том же процентном соотношении представлены тексты в производном корпусе Synek (SYNchronní Elektronický Korpus – составители в данном случае явно обыгрывают лексическое значение слова synek ‘сын, сынишка’). При этом корпус Synek достаточно велик, чтобы адекватно представить функционирование прилагательных в современном чешском письменном дискурсе. В частности, запрос на простые прилагательные дал 1 126 666 контекстов употребления 33 850 лексем. В других охватываемых проектом «Český Národní Korpus» электронных языковых корпусах, в том числе и в серии корпусов современных письменных чешских текстов SYN, соотношение стилей иное. Так, в корпусах SYN2005 и SYN2010 большую часть языкового материала составляет беллетристика (40%), публицистические тексты составляют 33%, специальная литература – 27%. Наконец, в корпусах SYN2015 и SYN2020 представлено равное соотношение художественной, специальной и публицистической литературы (33.3%). Важно заметить, что изменение жанрового состава языковых корпусов основано на периодическом

⁴ Český národní korpus. [Электронный ресурс]. Режим доступа: korpus.cz (Дата обращения 24.08.2025).

исследовании восприятия письменного языка: в то время как структура технической литературы по ее тематической ориентации изменялась лишь незначительно, структура журналистики, с другой стороны, претерпевала значительные изменения. Появление новых газет и журналов, цифровизация материалов и становление письменного дискурса в сетевом пространстве отразилось на соотношении разделов в корпусе и степени представленности той или иной литературы. Учитывая ограниченную сопоставимость данных частотности из корпусов с различным составом, мы работали с каждым корпусом отдельно. Тем не менее, среднее соотношение синтетических и аналитических вариантов было также рассчитано по совокупным данным всех корпусов, так как изменения в их составе отражают динамические характеристики чешского письменного дискурса 1990–2019 гг. Сведения о составе и объеме упомянутых корпусов приведены ниже (Таблица 1):

Таблица 1: Обзор состава и объема корпусов серии SYN

	Synek	SYN2000	SYN2005	SYN2010	SYN2015	SYN2020
Период, гг.	1990–1999	1990–1999	2000–2004	2005–2009	2010–2014	2015–2019
Объём, млн токенов	11,9	120,9	122,4	121,6	120,7	121,8
Состав, %	15% белл.	15% белл.	40% белл.	40% белл.	33% белл.	33% белл.
	25% спец.	25% спец.	27% спец.	27% спец.	33% спец.	33% спец.
	60% публ.	60% публ.	33% публ.	33% публ.	33% публ.	33% публ.
Простые прил., N контекстов	1 126 666	11 018 734	10 745 338	9 933 690	9 929 313	10 319 410
Простые прил., N лексем	33 850	43 059	56 567	49 883	55 444	54 563

Первоначально в качестве материала исследования был использован электронный корпус Synek объемом 11 959 431 токен, представляющий собой пропорционально редуцированный корпус современных письменных текстов SYN2000 объемом 120 908 724 токена, входящий в состав Чешского национального корпуса и подготовленный в 2000 г. на основе литературы периода 1990–1999 гг. Выбор уменьшенного корпуса был обусловлен потребностью

выработать процесс анализа контекстов для последующего применения с той же целью компьютерных инструментов: на корпусе меньшего объема было возможно произвести отладку фильтров и настройку анализирующих моделей.

Следующим этапом стало использование в качестве эмпирического материала больших корпусов серии SYN объемом около 120 млн токенов каждый. Так как объем данных корпусов приблизительно одинаковый, а их состав различен, было решено изучить результаты схожих запросов по каждому из корпусов. Затем, учитывая корреляцию состава корпуса с распространением тех или иных видов литературы в определенный период, полученные результаты предполагалось сопоставить и пронаблюдать развитие характеристик градационности качественного признака прилагательных за последние тридцать лет.

Методы исследования. Следует заметить, что обнаружение в корпусе контекстов, в которых употреблены аналитические формы сравнительной и превосходной степени чешских прилагательных, требовало последующего визуального контроля результатов на небольшой выборке, так как часть обнаруженных контекстов были выделены обслуживающей корпус программой ошибочно. Составители корпуса исходили из того, что опечатки и ошибки в письменном тексте заслуживают внимания как потенциальный объект корпусного анализа, и сознательно оставили во входящих в состав корпусов серии SYN текстах опечатки. В связи с этим были определены методы исследования и решения конкретных задач.

Для первичного отбора фактического материала использовался метод корпусного анализа: поиск в корпусе Synek осуществлялся с помощью корпусного менеджера Bonito (Manatee), версия 1.10.usnk. Для обработки собранного корпусного материала использовались системный и функциональный методы в том виде, в котором эти методы разработаны в рамках советской школы богемистики, основанной заслуженным профессором МГУ имени М.В. Ломоносова А.Г. Широковой. Построение таблиц осуществлялось с помощью статистического метода.

Обнаруженные в корпусе Synek опечатки и недоработки аннотации (например, отдельное написание приставки *nej-* и последующей части суперлатива, ошибочно обозначенной как компаратив) являлись стимулом для изменения ряда запросов с целью поиска аналогичных контекстов. Однако благодаря повышению точности аннотации в новых корпусах серии SYN число таких опечаток было невелико, и найденные контексты анализировались вручную посредством визуального контроля и относились к нужной категории.

При помощи изложенных методов были собраны данные в больших корпусах серии SYN. В качестве корпусного менеджера использовался веб-интерфейс Чешского национального корпуса. Однако ввиду технических ограничений функционала (отсутствия аннотации аналитических сочетаний) обработка данных корпусов потребовала применения вспомогательных методов, а именно моделирования контекстных закономерностей на основе выявленных контекстов и опечаток в корпусе Synek и обучения на них больших языковых моделей. Тем самым становилось возможным обработать массив данных из каждого сбалансированного корпуса серии SYN, превосходящий первичную выборку почти в двадцать раз; а суммарно – все контексты, стократно превышающие в объеме изначальную выборку.

Теоретической базой исследования для нас стали работы лингвистов Московской школы богемистики⁵, а также сотрудников отдела грамматики Института чешского языка Академии наук Чешской республики, в частности, выпускаемая ими «Большая академическая грамматика литературного чешского языка»⁶, а также издаваемые ими журналы «Слово и словесность»⁷, «Наша речь»⁸, «Корпус–грамматика–аксиология»⁹. В качестве *tertium comparationis*

⁵ Широкова А.Г. Методы, принципы и условия сопоставительного изучения грамматического строя генетически родственных славянских языков // Широкова А.Г., Васильева В.Ф., Изотов А.И., Ананьева Н.Е. Сопоставительные исследования грамматики и лексики русского и западнославянских языков. М.: Изд-во Моск. ун-та, 1998. С. 10–99.

⁶ Velká akademická gramatika spisovné češtiny. I., Morfologie. Druhy slov, tvoření slov / Fr. Štícha [et al.]. Praha : Academia, 2018. 1148 s.

⁷ Slovo a slovesnost. [Электронный ресурс]. Режим доступа: asjournals.lib.cas.cz/slovoaslovesnost/home (Дата обращения: 29.06.2025).

⁸ Naše řeč. [Электронный ресурс]. Режим доступа: asjournals.lib.cas.cz/naserec/home (Дата обращения: 29.06.2025).

полезными были предыдущий вариант академической грамматики чешского языка, а также университетская «Грамматика современного чешского языка»¹⁰. Для определения особенностей чешской градационности на общеславянском фоне дополнительно использовались труды Н.Д. Русинова, М.Л. Ремневой и А.М. Камчатнова по старославянскому языку, Й. Марвана, В. Барнета и В. Матезиуса – по чешскому языку.

Несмотря на развитие корпусных исследований на материале чешского языка, **степень разработанности темы** оставляет пространство для выявления характеристик чешских прилагательных. В частности, вопрос сосуществования аналитических и синтетических форм компаративов и суперлативов не раскрыт ни в «Русской грамматике»¹¹, ни в более современных изданиях «Большая академическая грамматика литературного чешского языка»¹² и «Грамматика современного чешского языка»¹³. Лишь в последнем издании утверждается о наличии форм обеих степеней сравнения у 3% лексем, однако детальной статистики не приводится. Во многом это обусловлено масштабом корпуса: с одной стороны, обширный материал позволяет исследовать спорные вопросы грамматики, с другой – требует вычислительных ресурсов для анализа. Стоит заметить, что в настоящем исследовании было проанализировано более 2.5 млн контекстов; обработка такого объема данных заняла бы в разы больше времени при использовании технологий, доступных 10–15 лет назад.

Научная новизна исследования заключается в синтезе методов советской школы богемистики и доступных на сегодняшний день компьютерных инструментов лингвистического исследования: электронных языковых корпусов, лемматизации, форматирования с применением регулярных выражений, фильтрации при помощи нейросетевых моделей и

⁹ Korpus-Gramatika-Axiologie. [Электронный ресурс]. Режим доступа: asjournals.lib.cas.cz/korpus-gramatika-axiologie/home (Дата обращения: 29.06.2025).

¹⁰ Mluvnice současné češtiny. 1, Jak se píše a jak se mluví / V. Cvrček [et al.]. Praha: Karolinum, 2010. 353 s.

¹¹ Русская грамматика / V. Barnetová, H. Běličová-Křížková, O. Leška, Z. Zkoumalová, V. Straková. Díl 1. Praha: Academia, 1979. 664 s.

¹² Velká akademická gramatika...

¹³ Mluvnice současné češtiny...

сценариев на языке программирования и обработки данных при помощи статистических модулей электронных таблиц.

Работа обладает **теоретической значимостью** для грамматики и стилистики чешского языка, так как освещает не изученный ранее вопрос о соотношении синтетических и аналитических средств образования компаративов и суперлативов прилагательных, а также об их функционировании в современном письменном дискурсе. Теоретическая значимость также обусловлена разработкой и улучшением методологии проведения корпусных исследований, где анализируются большие объемы данных (big data) при помощи компьютерных инструментов.

Практическая значимость заключается в применимости алгоритма поиска словоформ и анализа результатов для проведения дальнейших исследований в области сопоставительной славистики, а также для обучения большой языковой модели (large language model), которая затем сможет применить данный алгоритм для аналогичного анализа текстов, вышедших после 2020 г. Так как нами применялись современные компьютерные инструменты обработки лингвистического материала, результаты данного исследования находят применение в прикладных лингвистических дисциплинах, таких как автоматическая обработка естественного языка (natural language processing) и машинное обучение (machine learning), а также в академических общих курсах грамматики и стилистики чешского языка.

В перспективе исследование может быть выведено за пределы чешского языка и славянских языков вообще, поскольку Чешский национальный корпус базируется на синтаксисе CQL и платформе Sketch Engine, содержащей, кроме того, еще 836 корпусов для 108 языков мира. То есть приведенные в настоящем исследовании запросы могут быть адаптированы под другие языки с сохранением синтаксиса или применяться с частичной совместимостью на сторонних корпусных движках, как то: AntConc или BlackLab.

На основе проведенных изысканий нейросетевые модели формируют алгоритмы, не зависящие от изучаемого языка или не ограничивающиеся им.

Полученные таким образом результаты могут быть полезны для совершенствования систем машинного перевода, синтеза речи, а также использоваться в практических курсах по грамматике, стилистике и теории перевода.

На защиту выносятся следующие основные положения:

1. Чешская грамматическая норма допускает сосуществование аналитических и синтетических форм степеней сравнения прилагательных, образуемых от одинаковых позитивов.
2. В лексической системе чешского языка выделяется ряд прилагательных, образующих простые и сложные формы обеих степеней сравнения.
3. Синтаксис чешского языка допускает возможность выноса квантификатора степени за пределы синтагмы однородных членов.
4. При освоении чешским языком новой адъективной лексики доминирующим морфологическим механизмом выступает синтетическая модель формообразования компаратива и суперлатива.
5. По соотношению синтетических и аналитических средств выражения градационности качественного признака прилагательных чешский язык демонстрирует уникальное типологическое отклонение от тенденций, характерных для генетически родственных западно- и восточнославянских языков.

Личный вклад автора состоит в сборе, создании программного обеспечения для обработки корпусных данных, непосредственной обработке и анализе материала.

Степень достоверности полученных результатов подтверждена большим объемом исследованного материала, а также применением надежных методов исследования, описанных ранее. Научные положения и выводы в диссертации последовательно аргументированы и подкреплены фактическими данными.

Апробация. Результаты и выводы, полученные в процессе исследования, прошли апробацию на Международной научной конференции «Актуальные

проблемы славянской филологии, культуры и журналистики», проходившей 21–22 мая 2021 г. в Нижегородском государственном университете имени Н.В. Лобачевского.

Основные положения, результаты и выводы диссертации отражены в 7 статьях общим объемом 4,03 п.л., 4 из которых – объемом 2,77 п.л. – опубликованы в рецензируемых научных изданиях, рекомендованных для защиты в диссертационном совете МГУ по специальности 5.9.6 Языки народов зарубежных стран (славянские языки).

Структура работы. Диссертация состоит из введения, четырех глав, заключения, библиографии – списка источников фактического материала, языковых корпусов, а также списка использованной и цитируемой литературы, включающего 160 наименований.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **Введении** обосновывается актуальность работы, определяется предмет и объект диссертационного исследования, раскрываются его цели и задачи, характеризуются методы исследования и использованный фактический материал, научная новизна, теоретическая и практическая значимость, формулируются основные выносимые на защиту положения. Также во Введении представляются сведения об объеме и составе языковых корпусов, по которым далее проводится корпусный анализ.

В **Главе 1** «Теоретические основы исследования» приводятся основные сведения о развитии изучения градуационности и парадигм степеней сравнения. В данной главе дается определение градуационности по Ю.Л. Воротникову¹⁴ как «способности качественного признака выступать в одном носителе в той или иной степени своего проявления по сравнению с нормой или с другим носителем». При этом градуационность присуща не только прилагательным, а,

¹⁴Воротников Ю.Л. Структура функционально-грамматической сферы градуационности качественного признака // Научные доклады высшей школы. Филологические науки. 1987. № 4. С. 78–81.

например, существительным и глаголам, ср.: рус. *кот, котенок*, чеш. *dát* «дать», *dávat* «давать». У прилагательных же явление градационности представлено не только парадигмой степеней сравнения, но и разнообразием форм категории субъективной оценки, ср.: рус. *большенький, большущий*, чеш. *veledůležitý* «важнейший», *široširý* «широчайший». В контексте этого упоминаются труды Л.Г. Яцкевич в области изучения и построения восьмиступенчатой парадигмы градационности¹⁵, проводится параллель между парадигмой градационности русских прилагательных и чешских, определяются модели для анализа синтетичности форм категории степеней сравнения. В частности, это синтетические модели, образуемые суффиксами *ějš/ejš/š* и приставкой *nej-*, а также аналитические модели, образуемые при помощи квантификаторов *víc/více/silněji* и *nejvíc/nejvíce/nejsilněji*.

В главе приводятся выдержки из грамматик чешского языка разных лет, чтобы представить развитие общеславянских моделей формирования степеней сравнения в современном чешском языке. Так, взгляд на формы степеней сравнения прилагательных меняется за последние полвека.

В «Русской грамматике», изданной в Праге в 1979 г., отмечается, что в современном чешском языке «степени сравнения прилагательных представлены системой простых форм», а в случаях, когда «по формально-техническим причинам [чешское] прилагательное не образует простых форм сравнительной и/или превосходной степени», возможно образование «описательных сочетаний *více / nejvíce* (или же *silněji / nejsilněji*) + положительная степень: *více/nejvíce překvapující*»¹⁶.

В более новой «Грамматике современного чешского языка» 2010 г. сосуществование простых и сложных форм, восходящих к одному позитиву, не исключается. Однако указывается, что формы сравнительной или превосходной степени образуют около 6 % прилагательных, при этом формы всех трех

¹⁵ Яцкевич Л.Г. Словообразовательная категория степени интенсивности признака в русских народных говорах // Лексический атлас русских народных говоров (Материалы и исследования) 2014. Санкт-Петербург: Нестор-история, 2014. С. 700–733.

¹⁶ Русская грамматика... С. 347–350.

степеней в современном чешском языке представлены лишь у 3 % чешских прилагательных, что составляет примерно две тысячи лексем¹⁷. При этом не указано, как распределяются между собой синтетические и аналитические модели, какое реальное соотношение между ними существует в тех случаях, если прилагательное может формировать степени сравнения по обеим моделям.

При беглом рассмотрении парадигмы степеней сравнения в чешском и русском языках, где модели формирования степеней сравнения восходят к общеславянским архетипам **jъs*, **ejъs*, в чешском языке заметно преобладание синтетического способа выражения значения сравнения. Например, если в русском языке лексема *высокий* (154329 вхождений) продуктивно образует аналитические сочетания *более высокий* (5311 вхождений), *наиболее высокий* (448 вхождений), *самый высокий* (4521 вхождение) и синтетические формы *высший* (61 879 вхождений) и *выше* (19028 вхождений)¹⁸, то в чешском языке лексема *vysoký*, встречается 22606 раз в форме *vyšší* и 15276 раз в форме *nejvyšší*, но аналитических сочетаний типа **(nej)víc(e) vysoký* в корпусе не обнаружено¹⁹.

С другой стороны, в Чешском национальном корпусе встречаются сочетания вида *více hrdý* «более гордый», *více závislý* «более зависимый» с качественными прилагательными, в норме образующими компаративы и суперлативы синтетически. То есть данные корпуса подвергают сомнению утверждение авторов пражской «Русской грамматики» об образовании в чешском языке аналитических форм компаративов и суперлативов исключительно при невозможности образования форм синтетических. Таким образом, выдвигается гипотеза о наличии конкуренции простых и сложных форм степеней сравнения прилагательных, которую предполагается проверить на материале сбалансированных языковых корпусов серии SYN.

В **Главе 2** «Опыт корпусного исследования суперлативов» освещается процесс корпусного исследования и анализа контекстов на примере суперлативных форм. Объясняется необходимость автоматизации анализа при

¹⁷ Mluvnice současné češtiny... С. 205

¹⁸ Данные частотности по основному корпусу Национального корпуса русского языка.

¹⁹ Данные частотности по корпусу SYN2020.

помощи компьютерных инструментов лингвистического исследования, приводятся технические сведения об используемых программах. Также приводятся запросы для поиска контекстов суперлативов в корпусах:

1. [tag="AA.....1.*"] – простые прилагательные.
2. [(tag="AA.....3.*") & (word="[Nn][Ee][Jj].*")] – синтетические суперлативы, напр. *nejnovější*.
3. ([word="[Nn][Ee][Jj][Vv][Íí][Cc][Ee]"] [tag="AA..[134567]....1.*"]) – сочетания *nejvíce* с прилагательными.
4. ([word="[Nn][Ee][Jj][Vv][Íí][Cc]"] [tag="AA..[134567]....1.*"]) – сочетания *nejvíc* с прилагательными.
5. ([word="[Nn][Ee][Jj][Vv][Íí][Cc][Ee]?"] [tag="AA..[2].....*"]) – сочетания *nejvíc/nejvíce* с прилагательными в форме генитива.
6. ([word="[Cc][Oo]"] [word="[Nn][Ee][Jj][Vv][Íí][Cc][Ee]?"] [tag="AA..[134567]....1.*"]) – неколичественные сочетания *nejvíc* или *nejvíce* с прилагательными.
7. ([word="[Nn][Ee][Jj][Ss][Ii][Ll][Nn][Ěě][Jj][Ii]"] [tag="AA.....1.*"]) – сочетания *nejsilněji* с прилагательными.
8. ([word="[Nn][Ee][Jj]"] [tag="AA.....2.*"]) – ошибочные сочетания префикса *nej-* с прилагательными, маркированными как компаратив (фильтр отдельного написания простых суперлативов).
9. ([word="[Nn][Ee][Jj][Vv][Íí][Cc][Ee]?"] [word="[Nn][Ee][Jj][Ss][Ii][Ll][Nn][Ěě][Jj][Ii]"] [tag="AA..[134567]....3.*"]) – плеонастические сочетания *nejvíc*, *nejvíce* или *nejsilněji* с простыми суперлативами.
10. ([word="[Nn][Ee][Jj][Vv][Íí][Cc][Ee]?"] [tag="AA..[134567].*"] [word="[AaIi]?"] [tag="AA..[134567].*"]) – неколичественные сочетания *nejvíc* или *nejvíce* с однородными прилагательными, связанными союзами *a* или *i*.
11. ([tag="AA..[134567]....3.*"] [word="[AaIi]?"] [tag="AA..[134567]....3.*"]) – неколичественные сочетания двух однородных суперлативов, связанных союзами *a* или *i*.

Приведенные запросы позволяют получить из корпуса контексты употребления адъективных форм превосходной степени. Однако не все полученные контексты релевантны: они могут содержать опечатки²⁰, ошибки разметки, дубликаты²¹ и омографичные формы.

Например, в корпусе Synek запрос на сочетания «*nejvíce* + положительная степень» ([word="[Nn]ejvíce" [tag="AA.....1.*"]]²² дал 162 контекста, в 57 из которых квантификатор *nejvíce* не относился к прилагательному, ср.: <*Nejvíce správných*> *odpovědi zaznamenala agentura...* «Больше всего правильных ответов зафиксировала агентура...», *vložit do systému co <nejvíce české> práce* «вложить в систему как можно больше чешской работы», *v těchto profesích je však také <nejvíce nabízených> volných míst* «этим профессиям соответствует больше всего предлагаемых вакантных мест» и т. п.

А помимо суперлатива *nejlepší*, соотносимого с позитивом *dobry* «хороший», в том же корпусе обнаружены контексты с формами *nejlepšíší*, *nejlepčím*, *nejlepšíííí*, *nejlep-ším*, *nejlepšách* – и все эти формы представлены в автоматически сгенерированной таблице отдельными строками. Требовалось внести в таблицу правки и верно соотнести такие формы, суммировав число контекстов для исходной леммы. Таким образом, запросы необходимо было дорабатывать, а результаты – фильтровать.

После фильтрации полученных контекстов, их семантического анализа, сбора и консолидации информации на материале небольшой выборки контекстов корпуса Synek выявляются недействительные контексты (с опечатками, неверно распознанными суперлативами, неправильно

²⁰ Создатели Чешского национального корпуса сознательно оставляли существующие в текстах ошибки, считая их потенциальным объектом корпусного анализа.

²¹ Материалы корпусов серии SYN охватывают период с 1990 по 2019 г., и в них присутствуют тексты, где одна и та же лексема представлена в современной и прежней орфографии, ср.: *exkluzivní*, *exkluzívni*. Реформа орфографии была проведена в 1993 г. и коснулась в частности правописания заимствованных слов.

²² Данный запрос использовался нами еще до выявления написаний целиком в верхнем регистре и внесения корректив в изначальный паттерн. Тем не менее, в корпусе Synek запрос ([word="[Nn][Ee][Jj][Vv][Íí][Cc][Ee]" [tag="AA.....1.*"]) также возвращал 162 контекста. Следует заметить, что в данном запросе не отфильтрованы коллокации с прилагательными в форме родительного падежа – то есть возможные сочетания с количественным значением.

соотнесенными с позитивом словоформами), из-за чего число обнаруженных аналитических сочетаний сократилось. Тем не менее, в текстах 1990–1999 гг., представленных в корпусе Synek, было найдено 108 случаев конкуренции между синтетическими и аналитическими суперлативами у 75 прилагательных, что составляет 6.8% от всего числа суперлативов, употребленных в простой форме. Составляется таблица конкуренции (Таблица 2).

Таблица 2: Активная конкуренция относительно частотных суперлативов по данным корпуса SYN2000 (первые 15 строк)

№	Позитив	Синтетический	N синт.	Аналитический	N аналит	% аналит.
1	postížený	nejpostíženější	70	nejvíc[e] postížený	132	65,35
2	zatížený	nejzatíženější	24	nejvíc[e] zatížený	18	42,86
3	podobný	nejpodobnější	27	nejvíc[e] podobný	16	37,21
4	patrný	nejpatrnější	76	nejvíc[e] patrný	26	25,49
5	ohrožený	nejohroženější	135	nejvíc[e] ohrožený	39	22,41
6	diskutovaný	nejdiskutovanější	100	nejvíc[e] diskutovaný	22	18,03
7	chutný	nejchutnější	61	nejvíc[e] chutný	12	16,44
8	exponovaný	nejexponovanější	42	nejvíc[e] exponovaný	8	16
9	problematický	nejproblematictější	102	nejvíc[e] problematický	14	12,07
10	obchodovaný	nejobchodovanější	214	nejvíc[e] obchodovaný	21	8,94
11	používaný	nejpoužívanější	210	nejvíc[e] používaný	20	8,7
12	žádaný	nejžádanější	391	nejvíc[e] žádaný	16	3,93
13	rozvinutý	nejrozvinutější	99	nejvíc[e] rozvinutý	4	3,88
14	sporný	nejspornější	77	nejvíc[e] sporný	3	3,75
15	sledovaný	nejsledovanější	391	nejvíc[e] sledovaný	14	3,46

Полученные данные используются для аналогичного анализа выборки из сбалансированных корпусов серии SYN, каждый из которых десятикратно превосходит корпус Synek по объему токенов. В данных корпусах также выявляется конкуренция синтетических и аналитических суперлативов. Консолидировав данные, мы выявили 901 лексему, встречающуюся в форме аналитического и синтетического суперлатива в период с 1990 по 2019 гг. Общая доля прилагательных, которые обнаруживают конкуренцию простых и сложных форм, составила в среднем 16.4% (Рисунок 1).

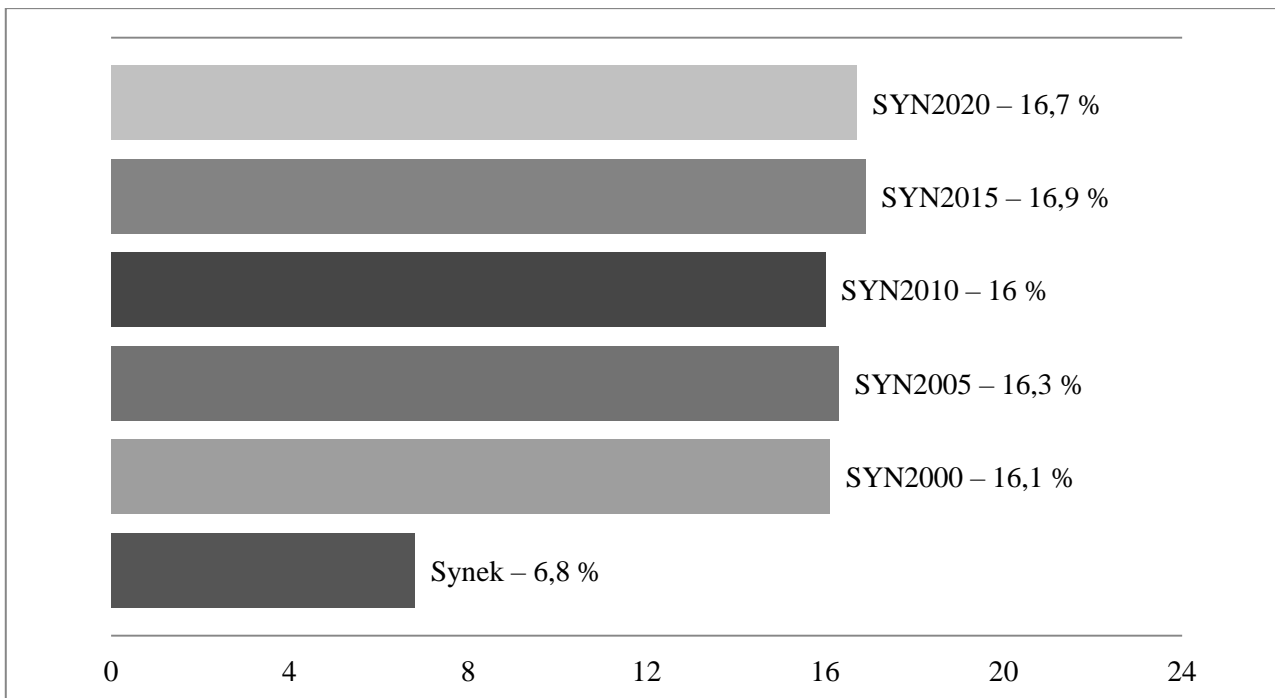


Рисунок 1: Доля лексем, обнаруживающих конкуренцию суперлативов в корпусах серии SYN

В Главе 3 «Опыт корпусного исследования компаративов» аналогичное исследование проводится на материале форм сравнительной степени. В главе приводятся запросы, по которым производится поиск и фильтрация контекстов:

1. [tag="AA.....2.*"] – синтетические компаративы, напр. *větší*.
2. ([word="[Vv][Íí][Cc][Ee]" [tag="AA..[134567]....1.*"]) – сочетания *více* с прилагательными.
3. ([word="[Vv][Íí][Cc]" [tag="AA..[134567]....1.*"]) – сочетания *víc* с прилагательными.
4. ([word="[Vv][Íí][Cc][Ee]?" [tag="AA..[2].....*"]) – сочетания *víc* или *více* с прилагательными в форме генитива.
5. ([word="[Ss][Ii][Ll][Nn][Ěě][Jj][Ii]" [tag="AA.....1.*"]) – сочетания *silněji* с прилагательными.
6. ([word="[Vv][Íí][Cc][Ee]?" [tag="AA..[134567]....[23].*"]) – сочетания *víc/více* с компаративами или суперлативами.
7. ([word="[Nn][Ee][Jj]" [tag="AA.....2.*"]) – фильтр суперлативных форм, ошибочно размеченных как сочетания *nej-* с компаративами.

8. ([word="[Vv][Íí][Cc][Ee]?"][tag="AA..[134567].*"][word="[AaIi]?"][tag="A A..[134567].*"]) – неколичественные сочетания *víc* или *více* с однородными прилагательными, связанными союзами *A* или *I*.
9. ([tag="AA..[134567]...2.*"][word="[AaIi]?"][tag="AA..[134567]...2.*"]) – неколичественные сочетания двух однородных компаративов, связанных союзами *a* или *i*.

Явление конкуренции у компаративных форм наблюдается чаще, чем у суперлативных, и воспроизводится на материале каждого из исследуемых языковых корпусов. Выявляется в среднем 21.5% прилагательных, обнаруживающих конкуренцию простых и сложных форм (Рисунок 2).

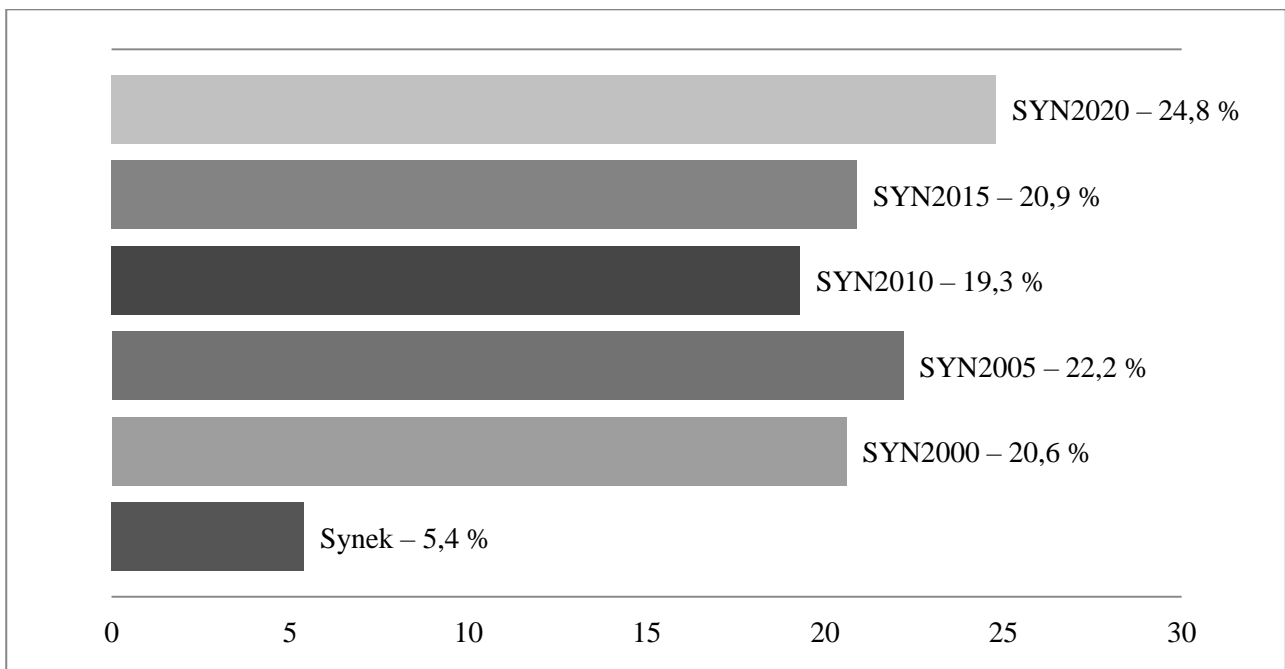


Рисунок 2: Доля лексем, обнаруживающих конкуренцию компаративов в корпусах серии SYN

В Главе 4 «Итоги и перспективы исследования» обобщаются сведения об анализе суперлативов и компаративов, а также выявляются такие лексемы, которые образуют полную парадигму степеней сравнения – то есть и синтетические, и аналитические формы обеих степеней. Список таких прилагательных приводится для каждого корпуса, после чего выводятся лексемы, которые обнаруживают конкуренцию на полной парадигме в течение последних тридцати лет – встречаются во всех взятых корпусах (Таблица 3).

Таблица 3: Прилагательные с полной парадигмой аналитических и синтетических форм степеней сравнения в чешском письменном дискурсе 1990–2019 гг. по данным сбалансированных корпусов серии SYN

№	Чешский позитив	Русский эквивалент	N синт.	N аналит.	% аналит.
1.	český	чешский	31	64	67,37
2.	postížený	пострадавший	311	398	56,14
3.	hrdý	гордый	52	66	55,93
4.	závislý	зависимый	357	234	39,59
5.	využívaný	используемый	167	98	36,98
6.	sociální	социальный	57	26	31,33
7.	patrný	заметный	467	188	28,7
8.	ohrožený	находящийся под угрозой	754	225	22,98
9.	podobný	похожий	581	172	22,84
10.	diskutovaný	обсуждаемый	327	83	20,24
11.	zvědavý	любопытный	134	31	18,79
12.	konkurenceschopný	конкурентоспособный	149	30	16,76
13.	znepokojivý	тревожный	180	33	15,49
14.	zřejmý	очевидный	468	50	9,65
15.	náchylný	склонный	772	80	9,39
16.	otevřený	открытый	1337	105	7,28
17.	lidský	человеческий	567	36	5,97
18.	zranitelný	уязвимый	927	54	5,5
19.	viditelný	видимый	1061	59	5,27
20.	rozvinutý	развитый	683	38	5,27
21.	problematický	проблематичный	908	50	5,22
22.	rizikový	рисковый	1059	54	4,85
23.	vyhledávaný	востребованный	832	40	4,59
24.	žádaný	желанный	1739	70	3,87
25.	přátelský	дружелюбный	514	18	3,38
26.	rozšířený	распространённый	3805	125	3,18
27.	spokojený	довольный	1397	41	2,85
28.	propracovaný	продуманный	786	21	2,6
29.	citlivý	чувствительный	2840	51	1,76
30.	pravděpodobný	вероятный	3871	69	1,75
31.	nápadný	заметный	1491	24	1,58
32.	populární	популярный	4534	48	1,05
33.	vzdálený	удалённый	6005	62	1,02
34.	šťastný	счастливый	5712	49	0,85
35.	oblíbený	популярный	9398	57	0,6
36.	známý	известный	13973	81	0,58
37.	efektivní	эффективный	4743	27	0,57
38.	náročný	требовательный	9858	35	0,35
39.	zajímavý	интересный	12124	37	0,3
40.	důležitý	важный	46462	37	0,08
			141433	3066	2,12

У 40 лексем, которые отражают последовательную конкуренцию простых и сложных форм степеней сравнения в чешском письменном дискурсе 1990–2019 гг., доля аналитических употреблений составляет лишь 2.12% (Рисунок 3). Это означает преимущественное употребление синтетических форм, даже когда аналитические существуют и периодически фигурируют в дискурсе.

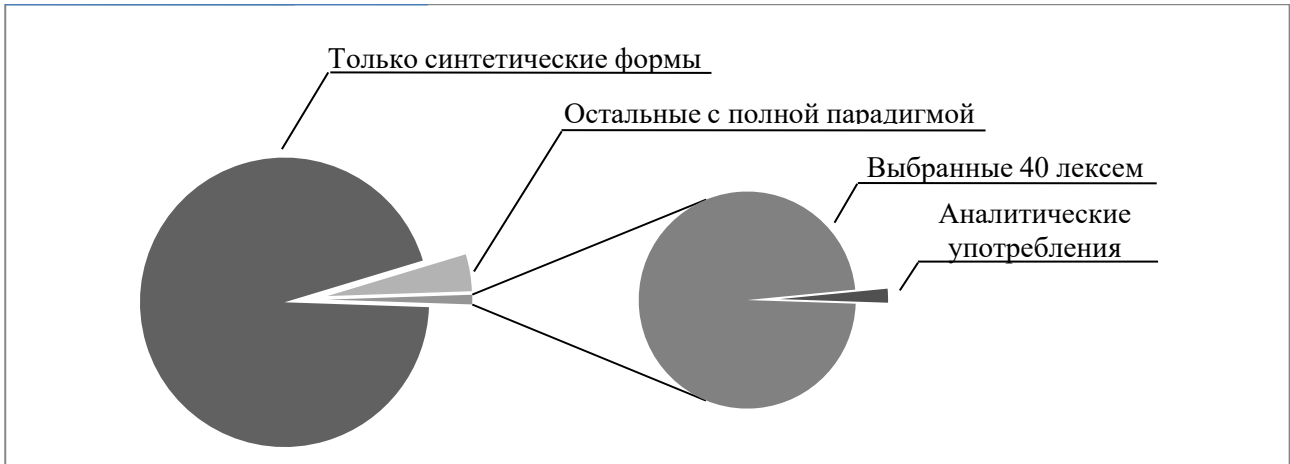


Рисунок 3: Средняя доля аналитических употреблений у 40 лексем, обнаруживающих конкуренцию в корпусах SYN2000-SYN2020, а также доля таких лексем среди всех "синтетических" прилагательных

Для первичного определения специфики градационности качественного признака чешских прилагательных приводится сравнение показателей синтетичности для чешского, русского и польского языков²³ – на основе анализа десяти наиболее частотных качественных прилагательных (Таблица 4).

Таблица 4: Доли аналитических сочетаний для десяти наиболее частотных качественных прилагательных в польском, русском и чешском языках

№	Польский		Русский		Чешский	
	Позитив,	%	Позитив,	%	Позитив,	%
	польский	аналит.	русский	аналит.	чешский	аналит.
1	dobry	0,1	новый	16,45	velký	0
2	nowy	0,19	большой	10,18	nový	0,11
3	mały	0,06	старый	13,78	dobrý	0
4	duży	0,02	молодой	32,02	mały	0
5	stary	0,16	белый	20,26	vysoký	0
6	piękny	0,24	полный	33,39	starý	0,01
7	wysoki	0,04	хороший	0,76	mładý	0,02
8	długi	0,12	чёрный	48,72	dlouhý	0
9	zły	0,43	маленький	7,73	důležitý	0,12
10	zimny	5,53	высокий	23,1	špatný	0,01
		0,15		9,232		0,011

²³ На материалах SYN2020, основного корпуса НКРЯ и сбалансированного подкорпуса НКЯР.

ЗАКЛЮЧЕНИЕ

В **Заключении** приводятся выводы по достигнутой цели исследования и кратко освещаются перспективы дальнейшей разработки темы диссертации и применения полученных результатов в последующих исследованиях.

Проведенный анализ позволил прийти к следующим выводам по цели исследования:

1. Вопреки данным пражской «Русской грамматики», некоторые чешские прилагательные образуют аналитические компаративы и суперлативы даже при возможности использовать синтетические формы. В среднем конкуренция наблюдается в 16.4% случаев у суперлативов и в 21.5% случаев у компаративов.
2. Полной парадигмой синтетических и аналитических суперлативов и компаративов обладает примерно 1% прилагательных, из которых 40 лексем встречаются во всех исследуемых корпусах чешского письменного дискурса на протяжении последних тридцати лет.
3. В 1% случаев вместо сочетания двух простых компаративов или суперлативов используется общий квантификатор степени для двух однородных позитивов.
4. На примере категории степеней сравнения показано, что градационности качественного признака чешских прилагательных присуща высокая степень синтетизма. С большой долей вероятности прилагательное, входящее в чешский письменный дискурс, будет образовывать формы степеней сравнения по синтетической модели.
5. На общеславянском фоне чешский язык выделяется развитым синтетизмом, отличаясь в аспекте градационности качественного признака прилагательных в десятки раз от польского языка и в сотни раз – от русского.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ДИССЕРТАЦИИ

Научные статьи, опубликованные в изданиях, рекомендованных для защиты в диссертационном совете Московского государственного университета имени М.В. Ломоносова по специальности 5.9.6. Языки народов зарубежных стран (славянские языки)

1. *Изотов А.И., Морозов Д.А.* Степени сравнения прилагательных в современном чешском письменном дискурсе // Вестник Московского университета. Серия 9. Филология. 2025. № 5. С. 85–95. EDN: КМКРVI. Импакт-фактор 0,288 (РИНЦ). 0,29 п.л. из 0,58 п.л.
2. *Изотов А.И., Морозов Д.А.* Суперлатив в современном чешском письменном дискурсе: опыт корпусного анализа // Славянский альманах. 2025. № 3–4. С. 130–148. EDN: CVDXIS. Импакт-фактор 0,322 (РИНЦ). 0,55 п.л. из 1,09 п.л.
3. *Морозов Д.А.* Опыт корпусного анализа компаративов в современном чешском письменном дискурсе // Litera. 2025. № 7. С. 66–76. EDN: АТСННJ. Импакт-фактор 0,333 (РИНЦ). 0,58 п.л.
4. *Морозов Д.А.* Сопоставительный анализ синтетичности форм степеней сравнения в чешском и русском языках // Litera. 2025. № 8. С. 180–189. EDN: ОКОНKS. Импакт-фактор 0,333 (РИНЦ). 0,52 п.л.

Научные статьи, опубликованные в других изданиях

5. *Изотов А.И., Морозов Д.А.* Опыт корпусного анализа конкуренции синтетических и аналитических форм сравнительной и превосходной степеней сравнения прилагательных (на материале современного чешского языка) // Филологические науки. Вопросы теории и практики. 2025. Т. 18. № 8. С. 3287–3293. EDN: RLBMWO. Импакт-фактор 0,444 (РИНЦ). 0,2 п.л. из 0,4 п.л.
6. *Изотов А.И., Морозов Д.А.* Опыт корпусного анализа степеней сравнения прилагательных (на материале современного чешского языка) //

Филологические науки. Вопросы теории и практики. 2025. Т. 18. № 7. С. 2677–2685. EDN: KMDBMB. Импакт-фактор 0,444 (РИНЦ). 0,23 п.л. из 0,46 п.л.

7. *Морозов Д.А., Рылов С.А.* Функционально-грамматическая градационность качественного признака в современных славянских языках: универсальное и локальное // Актуальные проблемы славянской филологии, культуры и журналистики: Материалы I Международной научной конференции. Нижний Новгород: ННГУ им. Н.И. Лобачевского, 2021. С. 57–63. EDN: QXMDHF. Импакт-фактор отсутствует. 0,35 п.л. из 0,4 п.л.