




Прикладная лингвистика Applied Linguistics

Научная статья

УДК 8'33

DOI: <https://doi.org/10.18721/JHSS.12404>

КОРПУСНЫЕ ТЕХНОЛОГИИ В ИССЛЕДОВАНИИ И ОБУЧЕНИИ: МЕТОДИКА АНАЛИЗА ИМЕННЫХ ГРУПП В НАУЧНОМ ТЕКСТЕ

Л.Н. Беляева¹ , О.Н. Камшилова^{1,2} ,
М.Л. Малаховская¹ 

¹ Российский государственный педагогический университет
им. А.И. Герцена, Санкт-Петербург, Российская Федерация;

² Санкт-Петербургский университет технологий управления и экономики,
Санкт-Петербург, Российская Федерация

✉ onkamshilova@gmail.com

Аннотация. Современные методы прикладной лингвистики во многом определяются возможностями инструментов обработки текста, доступных не только лингвисту-исследователю, но и широкому кругу пользователей. К такого рода инструментам относятся корпусные менеджеры – комплекты инструментов, обслуживающие корпус текстов, встроенные или автономные. В статье на примере анализа английских многокомпонентных именных словосочетаний (именных групп – ИГ), важнейшего компонента и главного носителя терминологии в научных и научно-технических текстах, доказываем, что инструментарий корпусной лингвистики может быть с равным успехом применен как к исследованию структуры ИГ, так и к обучению грамматике ИГ, стратегии их понимания и построения, что представляется важным в аспекте овладения академическим и научным стилем. Основное внимание в статье уделяется многокомпонентным ИГ, построенным по модели существительное + существительное. На материале оригинальных исследовательских корпусов текстов разных предметных областей описываются их характеристики: частота, структура, синтаксические связи, способы трансформации в пределах текста (усложнение и свертывание), определяющая роль в создании синтаксической сложности научного текста. На основе современных методик, опирающихся на корпусные технологии, а также личного опыта авторов в статье предлагается процедура работы с ИГ в специализированном корпусе текстов, построенном в соответствии с предметной областью и конкретными интересами студентов/начинающих авторов, которая помогает идентифицировать, «распаковывать» и создавать подобные конструкции в научных текстах своей предметной области.

Ключевые слова: именная группа (ИГ), многокомпонентное именное словосочетание, научный текст, корпус текстов, корпусные технологии, корпусные инструменты.

Для цитирования: Беляева Л.Н., Камшилова О.Н., Малаховская М.Л. Корпусные технологии в исследовании и обучении: методика анализа именных групп в научном тексте // Общество. Коммуникация. Образование. 2021. Т. 12. № 4. С. 46–59. DOI: 10.18721/JHSS.12404

Статья открытого доступа, распространяемая по лицензии CC BY-NC 4.0 (<https://creativecommons.org/licenses/by-nc/4.0/>).

Scientific article

DOI: <https://doi.org/10.18721/JHSS.12404>

CORPUS TECHNOLOGIES IN RESEARCH AND EDUCATION: ANALYSING ENGLISH NOUN PHRASES IN SCIENTIFIC TEXTS

L.N. Beliaeva¹ , O.N. Kamshilova^{1,2} ,
M.L. Malakhovskaya¹ 

¹ Herzen State Pedagogical University of Russia,
St. Petersburg, Russian Federation;

² St. Petersburg University of Management Technologies and Economics,
St. Petersburg, Russian Federation

✉ onkamshilova@gmail.com

Abstract. Methods of applied linguistics today are largely determined by text processing tools available not only to researchers, but also to a wide range of users. The article demonstrates that text corpora and corpus tools may be well applied to both research and teaching of the structure and composition of such complicated units as multi-component noun phrases (NPs). Based on original text corpora, the research focuses on syntactic and structural characteristics and transformations of multi-component NPs, namely noun+noun constructions, within a text. The article also proves their role as markers of text complexity and scientific style. As the complexity of NPs requires special skills facilitating both reading and writing scientific texts, the article suggests a procedure of treating NPs in special text corpora, compiled by the learner or novice author of their own discipline texts, that helps to identify, “unpack” and compose NPs. The method proposed is based on some published practice and the authors’ personal experience in teaching students to deal with NPs in scientific and technical texts.

Keywords: noun phrase (NP), multi-component NP, scientific text, text corpus, corpus technologies.

Citation: L.N. Beliaeva, O.N. Kamshilova, M.L. Malakhovskaya, Corpus technologies in research and education: analysing english noun phrases in scientific texts, *Society. Communication. Education*, 12 (4) (2021) 46–59. DOI: 10.18721/JHSS.12404

This is an open access article under the CC BY-NC 4.0 license (<https://creativecommons.org/licenses/by-nc/4.0/>).

Введение. Постановка проблемы

Современные методы прикладной лингвистики во многом определяются возможностями инструментов обработки текста. Используемые сегодня программные средства становятся доступными не только лингвисту-исследователю, но и широкому кругу пользователей, владеющих общим алгоритмом работы в поисковых системах, благодаря их удобному интерфейсу и достаточно полной информационной (в т.ч. справочной) поддержке. К такого рода инструментам относятся корпусные менеджеры – комплекты инструментов, обслуживающие корпуса текстов, встроенные или автономные. Задача статьи – показать на примере анализа английских многокомпонентных именных словосочетаний (именных групп – ИГ), что инструментарий корпусной лингвистики может быть с равным успехом применен как к исследованию структуры ИГ, так и к обучению грамматике ИГ, стратегии их понимания и построения, что представляется важным в аспекте овладения академическим и научным стилем.

С тех пор как особая роль английского языка как языка международного общения была признана в научном и академическом сообществе, расширился поток учебной литературы по английскому языку для специальных (ESP) и образовательных (EAP) целей, а также для целей публикации результатов исследований (English for research publication purposes – ERPP). Изучение и использование английского языка в этих специальных сферах определили возрастающий интерес исследователей к процессу овладения языком теми, для кого он не является родным, с точки

зрения формирования академической грамотности. Академическая грамотность предполагает реализацию на двух параллельных, но связанных уровнях, которые были определены в [1] как дискурсивная грамотность и языковая грамотность. Под дискурсивной грамотностью понимается умение писать научную прозу, следуя логическим, структурным и стилистическим законам, определяемым стандартами соответствующей научной области, жанром и типом текста. Языковая грамотность подразумевает умелое использование в этих целях лексических и грамматических языковых средств. Обучение дискурсивной грамотности является обязательным, особенно для учебных и научных сообществ, чья национальная научная риторика отличается от современных международных дискурсивных норм [2]. Однако установлено, что у авторов научных текстов, для которых английский язык не является родным, возникает значительно больше проблем в овладении именно языковой грамотностью, они демонстрируют низкие лексические и грамматические навыки, специфические именно для научного текста [1, 3, 4]. Эти результаты нельзя игнорировать при обучении английскому языку для отмеченных выше целей, и языковой компонент учебной программы должен включать изучение лексических и грамматических структур научного текста.

Одной из таких особых структур и является именная группа (ИГ), синтаксическая структура с несколькими модификаторами в препозиции к именному ядру, часто исключительно именными, моделью которой является формула $N_1 + N_2 + \dots + N_n$ [5, 6]. Эта структура, характерная для текстов научного регистра, должным образом не рассматривается при обучении английскому языку. В результате даже авторы, владеющие языком на высоком уровне, делают в ее употреблении ошибки, которые существенно влияют на восприятие их текстов читателями, в том числе рецензентами международных научных журналов. Сложность английской ИГ, главного носителя терминологии в научных и научно-технических текстах, требует выработки специальных навыков для чтения и создания научного текста [7–10].

Следует иметь в виду, что и тексты переводов, и тексты, написанные на языке, который не является родным, по различным лингвистическим параметрам всегда отличаются от текстов, написанных носителями языка. Для подчеркивания этого отличия был введен термин *translationese* [11], под которым понимаются статистически значимые отличия, наблюдаемые между оригиналами и переводами [12–16]. С учетом этого рассмотрим пути создания корпусно-ориентированных педагогических технологий, которые можно использовать для минимизации ошибок в сфере употребления ИГ.

Несмотря на то, что ИГ является основным средством номинации научных объектов и понятий и ведущей/определяющей лексической формой в терминологии, по свойственным ей языковым/структурным характеристикам она должна быть отнесена, скорее, к сфере синтаксиса, а не лексики, а ее синтаксическая организация зависит от текста в целом. Поэтому установить способы взаимодействия многокомпонентных ИГ в текстах означает понять связи внутри самой ИГ. А эту задачу можно решить на основе корпусных технологий.

Методология и методика исследования

В основе исследования лежит корпусная методология. Обсуждаемые в статье наблюдения были сделаны на основе результатов исследований в оригинальных авторских корпусах специальных исследовательских текстов по различным предметным областям (корпус «Сейсмическая защита», объемом 1 миллион слов, корпус «Web- и лингвистические технологии», объем 3 468 000 слов)¹.

Методология исследования основывается на понимании ИГ как структурной единицы, которая, будучи функционально равной слову, фактически представляет собой сжатие, свертку предложения, что делает ИГ скорее единицами синтаксиса, а не словаря. Внутренняя структура

¹ Данные корпуса созданы авторами статьи в рамках исследований Центра теоретических и прикладных компьютерных исследований в филологии РГПУ им. А.И. Герцена.



именной группы должна коррелировать с внутренней структурой предложения, что и позволяет выявить синтаксические зависимости между компонентами словосочетания.

Анализ данных корпуса, полученных с помощью процедур построения частотных списков, n-грамм, кластеров и конкордансов, помогает определить роль текста в интерпретации синтаксических отношений в рамках ИГ и их значения, и, более того, доказывает, что корпусные методы полезны при обучении английскому языку для специальных и академических целей при интерпретации и создании ИГ.

Интерпретация ИГ и установление процедуры анализа структуры ИГ также включает такие методы, как синтаксическое моделирование, анализ контекста.

Результаты исследования

Именные многокомпонентные словосочетания – феномен, привлекавший внимание специалистов в самых разных областях лингвистики. Они становились объектом исследования психолингвистики, когнитивной семантики, переводоведения и, конечно, не раз морфологии и синтаксиса [17]. Рассмотрим основные проблемы, наиболее релевантные для изучения этой структуры в учебном курсе английского языка специальных и академических целей.

1. Корпусные исследования характеристик именных групп в научных текстах: частота, структура, внутренние связи

В текстах научного стиля частота ИГ чрезвычайно высока, при этом 30% всех препозитивных определений представлено существительными, что значительно превосходит долю их употребления в этой функции в текстах большинства других типов [18, с. 589]. Причиной высокой частоты ИГ в научных текстах считается компактность этой структуры, что ведет к тенденции использования тех языковых средств, которые дают авторам возможность выражать свои мысли наиболее экономно, а читателям – тратить меньше времени на чтение [19, 20]. Наши исследования в корпусах научных текстов показали, что наиболее частотными в английском языке являются двухкомпонентные ИГ, следующими по убыванию частоты в научных и научно-технических текстах являются ИГ с тремя элементами, но их частота втрое ниже [6].

Внутренняя структура именной группы должна коррелировать с внутренней структурой предложения, что и позволяет выявить синтаксические зависимости между компонентами словосочетания. Вопрос в том, как распознать эту структуру в сжатой форме ИГ, поскольку маркеры связей между реальными компонентами и типы связей между ними, которые обычно показаны в предложении с помощью различных средств, в английской именной группе отсутствуют. Сжатие структуры предложения, внешнее упрощение структуры и формы ИГ вызывают ее семантическое усложнение [6, с. 2907].

Существуют два способа введения ИГ в реальный текст: путем добавления лексемы-определителя к ранее используемой или терминологической ИГ, создавая новую, более сложную номинацию: *machine translation* => *machine translation method*, *machine translation service*, *machine translation program*, или путем устранения повторяющихся единиц, сворачивая таким образом структуру предложения до многокомпонентной ИГ: *syntactic dependency*, *syntactic formalism*, *syntactic dependency tree annotation* => *dependency annotation formalism* [21].

Первый способ представляет собой процесс постепенного усложнения, добавления конкретных характеристик ядру, второй – процесс последовательного свертывания, реализуемого на трех уровнях:

Уровень 1: переход от сложной именной группы (с предлогом) к простой путем инверсии элементов: *phrase-structure trees from dependency annotations* => *syntactic phrase-structure dependency trees annotation*;

Уровень 2: устранение дублирования компонентов в новой ИГ *syntactic phrase-structure dependency trees annotation* => *syntactic dependency tree annotation*.

Уровень 3: согласование сем и устранение компонентов с дублированными семами *syntactic dependency tree annotation* => *dependency tree annotation* [6:2908].

Итак, статус ИГ предполагает, что понимание ее структуры и значения требует экспликации связей в рамках текста. Анализ текстов в различных предметных областях показал, что если многокомпонентная ИГ появляется в тексте, она обычно сопровождается ИГ с 2 компонентами в ближайшем контексте, в пределах 2-3 предложений или ее можно найти в названии, перечне ключевых слов или аннотации (см. пример согласования названия и первого предложения в аннотации:

*Title: Shareable Subentries in Lexonomy as a Solution to the Problem of **Multiword Item Placement***

*Abstract: This paper introduces a new way of dealing with phraseology in dictionaries. A classical question in lexicography is whether **multiword items** such as ... (Proceedings of the XVIII EURALEX International Congress, p. 223)*

Анализ многокомпонентных ИГ в тексте ведет к установлению новых единиц на основе уже встретившихся в тексте и к распознаванию структуры предложения в новой ИГ, формируемой следующим образом:

- ИГ в роли постмодификатора включена в первую ИГ как преомодификатор:

method of machine translation ⇒ *machine translation method*,

service of machine translation ⇒ *machine translation service*

- семантика одного элемента из встретившейся в тексте двухкомпонентной ИГ входит в значение новой ИГ с семами других компонентов, результатом является новая ИГ с тремя компонентами, напр.:

cross-entropy method + *cross-entropy difference* ⇒ *cross-entropy difference algorithm*

- семантика одного из элементов включается как часть новой ИГ за счет экстралингвистической информации предметной области, напр.:

dependency relation + *dependency grammar analysis* ⇒ *dependency relation analysis*

которая в тексте может быть частью более длинной ИГ:

long distance dependency relation analysis.

Рассмотренные выше случаи трансформации именных групп, обусловленные когезией и когерентностью текста, не демонстрируют все возможные варианты, однако дают основу для понимания имплицитной внутренней структуры ИГ. Кроме того, именно двухкомпонентные ИГ представляют особые трудности для анализа и интерпретации.

Инструменты корпусного анализа дают возможность понимания внутренней структуры ИГ, а их доступность, относительная простота и возможность извлекать достоверные знания об устройстве этих сложных синтаксических структур могут быть полезными в овладении грамматикой английских ИГ. Рассмотрим, с какими сложностями сталкиваются иноязычные авторы, приступающие к написанию научных текстов на английском языке.

2. ИГ как маркер синтаксической сложности научного текста и показатель академической грамотности: теоретические основания к применению корпусного метода в обучении английским ИГ

Более высокая частота ИГ в научных текстах, чем в текстах других типов, свидетельствует, что эта конструкция является одной из типологически существенных черт научного стиля, и поэтому обязательно должна изучаться в учебном курсе английского языка, что требует учета ее синтаксической и семантической сложности.

В создании синтаксически сложных научных текстов у ИГ особая роль. Во-первых, в рамках большинства научных дисциплин использование существительного в качестве препозитивного определителя – один из распространенных способов производства новых терминологических единиц (например, *climate change*, *sex education*, *identity formation*), поэтому умение корректно



использовать эту структуру свидетельствует о хорошем знакомстве автора с предметной областью. Во-вторых, семантические отношения между компонентами этих единиц далеко не всегда эксплицитны. Понимание значения конкретного словосочетания часто определяется пониманием всего текста, что, в свою очередь, определяется референтной структурой конкретной предметной области. Следовательно, умелое использование именных групп свидетельствует, что автор не только хорошо владеет английским языком, но и достаточно продвинулся в овладении своей профессией [22].

В результате корпусных исследований текстов студентов, для которых английский не является родным [23], выявлены грамматические особенности разговорной речи и текстов научных статей и установлено, что наиболее ярко отличие этих двух типов речи проявляется в способах выражения атрибутивной семантики: для разговорного стиля характерно использование определительных придаточных предложений, для научного – сложных именных словосочетаний. Исходя из этого, была предложена схема развития навыков академического письма у студентов L2. При этом предполагается, что на начальном этапе студенты чаще всего в качестве определений используют прилагательные и определительные придаточные предложения, а по мере совершенствования в научном письме используют более сложные синтаксические структуры.

Эти данные нашли подтверждение в многочисленных исследованиях, показавших, что тексты авторов, наиболее продвинутых в академическом письме, независимо от того, является ли английский язык для них родным или иностранным, характеризует гораздо более плотное использование сложных именных групп, чем тексты менее опытных в этом отношении авторов [24–26].

Синтаксическая сложность является одним из требований научного текста [26], она воспринимается как показатель высокого профессионализма автора и даже его причастности к научному сообществу. Поэтому научить создавать синтаксически сложные тексты – одна из важнейших задач, стоящих перед преподавателями.

Обучение грамматике ИГ, характерных для английского научного стиля представляет собой весьма непростую задачу из-за информационной плотности ИГ, разнообразия семантических и синтаксических отношений между их элементами и их потенциальной неоднозначности [22].

Различные типы неверного использования ИГ были исследованы на материале научных текстов исследователей и аспирантов РГПУ им. А. И. Герцена, чей уровень владения языком оценивался как B2-C1 [27]. Проанализированные тексты представляли собой конспекты курсов лекций на английском языке и проекты статей, предназначенных для публикации в международных научных журналах. Анализ выявил понимание авторами назначения ИГ в научном тексте, хотя их состав очень часто был некорректным и, в результате, недоступным пониманию читателей. Анализ ошибок показал, что российские обучающиеся злоупотребляют именными группами с существительным в функции определителя в тех случаях, когда есть производное прилагательное, или используют в этой функции существительные во множественном числе. Подобные ошибки характерны не только для носителей русского языка [28; 29, с. 81].

Обзор наиболее популярных курсов по обучению английскому языку для общих, специальных и академических целей выявил лапидарность, ограниченность или полное отсутствие в учебных пособиях разделов, посвященных грамматике многокомпонентных ИГ с именными компонентами. Грамматика субстантивных конструкций не преподается обучающимся, независимо от уровня владения языком. Средства выражения атрибутивных связей отрывочны, не системны и не сопоставимы [28], типы связи компонентов ИГ анализируются поверхностно.

Многие лингвисты утверждают, что невозможно строго отличить двухкомпонентные именные конструкции от свободных словосочетаний из двух существительных, и заявляют, что ИГ находятся в точке пересечения морфологии и синтаксиса [29, 20]. Это может объективно отражать проблемы лингвистической интерпретации ИГ, но с лингводидактической точки зрения это препятствие, которое ведет к необудительным рекомендациям.

Интерпретация семантических отношений между элементами ИГ является другим препятствием. Перечни семантических отношений между элементами ИГ отличаются по количеству и типам. Так, например, выделяют 15 типов семантических отношений между двумя существительными в двухкомпонентной именной группе: сочинительные (*leather coat*), цели (*war fund*), источника (*whale meat*), времени (*summer holiday*), места (*heart disease*) и т.д. [18]. Предлагаются также перечни или более короткие, или более длинные (см., например, [20]), либо утверждается, что такие перечни вообще не имеют смысла [30].

Еще одной трудностью в понимании грамматики ИГ является отсутствие морфологических элементов, которые могли бы отмечать синтаксическую зависимость между ядерным словом и модификаторами. Внешняя простота ИГ вводит в заблуждение и чаще всего их значение можно определить только по контексту и знаниям предметной области [20, 21, 31].

Поэтому обучение пониманию и построению английских ИГ должно стать вопросом серьезного дидактического внимания и быть частью учебного плана в преподавании английского языка для специальных и академических целей. Мы предполагаем, что при изучении состава и структуры ИГ и их реальной реализации в научном/академическом тексте не-носителями языка следует использовать корпусные технологии с учетом специфичности ИГ в различных предметных областях.

Исследования научно-технических текстов в последние два десятилетия обнаружили, что языковые характеристики научных текстов больше зависят от предметной области или дисциплины, чем от родного языка и культурного фона автора [32, 22]. Самая высокая частота характерна для текстов точных наук, а самая низкая частота является характеристикой наук гуманитарных. Кроме того, оказалось, что частота ИГ зависит от того, является ли дисциплина фундаментальной или прикладной. Тексты прикладных дисциплин содержат больше ИГ, чем тексты фундаментальных. Кроме того, ИГ в текстах точных наук представляют собой более длинные последовательности, чем в текстах гуманитарных наук.

В целом, подобные исследования позволяют предположить, что обучение использованию ИГ теми, кто изучает английский язык для специальных целей, может быть эффективнее в рамках их собственной дисциплины, а не только на основе общего дидактического подхода. Этот специфический для конкретных дисциплин подход к обучению может поддерживаться применением корпусных технологий, которые используются для работы в аудитории или индивидуально, с одной стороны, и как ресурс для новых дидактических материалов, с другой.

3. Корпусные технологии как поддержка в интерпретации и построении именных групп в научном тексте

Компьютеризованный подход подразумевает использование корпусов текстов и корпусного инструментария, которые доказали свою эффективность, помогая обучающимся обогащать свой словарь и улучшать грамматику и стиль собственных научных текстов. Опираясь на опубликованные рекомендации [33–36] и личный опыт [37, 38], предлагаем процедуру использования корпусных технологий для обучения пониманию связей в многокомпонентных ИГ, проверки корректности их композиции в создаваемых научных и научно-технических текстах.

В соответствии с утверждением, что понимание синтаксической организации (композиции) ИГ существенно определяется контекстом, при обучении следует опираться на корпуса текстов, как известные, так и оригинальные, построенные самими обучающимися и включающие тексты их предметной области.

Предлагаемая процедура основана на использовании компьютерных средств анализа текстов, таких как Sketch Engine², Lexical Tutor³ и др. а также прикладной программы AntConc⁴, работа с которой иллюстрирует наш подход в данной статье. Эта программа является бесплатным и

² <https://www.sketchengine.eu/>

³ <https://www.lextutor.ca/>

⁴ <https://www.laurenceanthony.net/software/antconc/>



надежным комплектом инструментальных средств с рядом полезных функций для автоматической обработки текста [39]. Надежность и важность этого комплекта инструментов подтверждена исследователями текста и преподавателями [33–36, 39, 40].

Выбор этого инструментария определяется рядом причин, в том числе его доступностью и надежностью, а также относительно малым объемом обрабатываемого корпуса. Функции обработки текста включают получение частотных списков, конкорданса, кластеров и n-грамм.

Анализ данных корпуса (n-грамм, кластеров и конкордансов) помогает определить роль текста в интерпретации синтаксических отношений в рамках ИГ и их значения, и, более того, доказывает, что корпусные методы полезны при обучении английскому языку для специальных и академических целей при интерпретации и создании ИГ.

Сбор данных. Обучающимся английскому языку для специальных целей может быть полезно искать корпуса текстов в Интернете, типа коллекции профессиональных корпусов RCPCE⁵, а также и национальных корпусов текстов и корпусов текстов, включенных в интерактивные инструменты анализа текста типа Sketch Engine. Например, коллекция RCPCE содержит корпусы текстов 39 предметных областей, которые могут быть обработаны рядом инструментальных средств, разрешающих поиск отдельных слов/фраз или двух или более связанных слов/фраз и составление конкорданса. Кроме того, можно добавлять тексты, которые наиболее отвечают академическим и научным интересам обучающихся. Работа с AntConc направлена исключительно на собственный выбор исследователей/обучающихся, т.е. сбор материала может быть ограничен конкретными предметными областями и включать статьи, написанные экспертами в этой области, что является более эффективным согласно недавним исследованиям [36].

Процедура работы по созданию корпуса текстов в своей предметной области начинается с этапа нормализации текста и приведения его к читаемому формату (см. [39, с. 96–103]). Инструменты программы позволяют обнаружить частотные существительные, претенденты на ядро или модификатор ИГ. Извлечение кластеров и n-грамм полезно для установления длины ИГ и позиции существительного (ядро/модификатор) и определяет их частоту в корпусе. Составление конкорданса дает возможность увидеть существительное в непосредственном контексте

Рассмотрим работу этого комплекса инструментов на примере слова *segmentation* (частота 6) как кандидата на компонент ИГ в рамках одного текста предметной области «Машинный перевод». Задав размер n-граммы с существительным *segmentation* (4), получаем 6 тетраграмм:

segmentation (tr#29) grapheme boundaries
segmentation for the xlyff
segmentation rules are used
segmentation rules exchange) – xml
segmentation rules have been
segmentation rules, if the

Анализ полученных тетраграмм для слова *segmentation* показывает, что оно используется в тексте как однокомпонентная ИГ с частотой 1, в 4 других случаях это слово является частью ИГ с определителями: *segmentation rules* (3) и *segmentation rules exchange* (1). Чтобы определить статус лексической единицы *segmentation (tr# 29 grapheme boundaries)*, а также статус других 4, следует обратиться к непосредственному контексту, что позволяет сделать инструмент конкорданс. Анализ контекста подтверждает, что слово *segmentation* является определителем существительного в *segmentation rules*, последний элемент этой ИГ в свою очередь является определителем для слова *exchange* в трехкомпонентной ИГ *segmentation rules exchange*. Чтобы принять решение по поводу статуса словосочетания (*tr# 29 grapheme boundaries*, строки конкорданса недостаточно, для этой цели необходимо использовать инструмент «Просмотр файла» (File View), демонстрирующий широкий контекст, фрагмент текста, который позволяет увидеть, что данная тетраграмма не

⁵ RCPCE Profession-specific Corpora (polyu.edu.hk)

является синтаксически осмысленной, поскольку ее граница, помеченная круглой скобкой, вводит синтаксически свободный элемент ((*tr# 29 grapheme boundaries*)).

Таким образом, исследователь или учащийся может найти реально и корректно используемые ИГ различного состава и длины в текстах своей конкретной дисциплины. Инструментальные средства корпуса обеспечивают надежные факты для создания многокомпонентных ИГ, дают модели их корректного использования и тем самым могут компенсировать недостаточность дидактических материалов. Анализ именных групп в рамках пространства текста ведет к установлению процедур создания новых ИГ на основе показанных в тексте и к распознаванию сжатой структуры предложения в краткой форме ИГ.

Заключение

Как было показано, английские многокомпонентные ИГ представляют проблему для начинающих авторов научных текстов, если английский не является их родным языком. Сложность в интерпретации и овладении грамматикой таких ИГ обусловлена тем, что сжатость, свертка внутренних связей между компонентами ИГ при отсутствии эксплицитных показателей приводит к их семантическому усложнению и потенциальной неоднозначности.

При этом, как доказывают наши и другие исследования, многокомпонентные ИГ являются неотъемлемой характеристикой научного и научно-технического текста, так как а) удовлетворяют требованию плотности, предъявляемому к текстовой информации в научной прозе; б) служат маркером синтаксической сложности научного текста и тем самым свидетельствуют о квалификации автора; в) являются основным средством номинации научных объектов и понятий и ведущей/определяющей лексической формой в научной терминологии. Особенностью многокомпонентных ИГ как номинативных единиц является то, что они представляют собой свертку, синтаксические отношения внутри которой могут быть определены контекстным окружением ее отдельных элементов в пределах конкретного текста.

В ходе исследований в оригинальных корпусах научных текстов разных предметных областей нами было установлено, что наиболее частой и наиболее сложной для интерпретации являются двухкомпонентные ИГ, а частота ИГ с тремя компонентами втрое ниже. При этом мы показали, что если многокомпонентная ИГ появляется в тексте, она обычно сопровождается ИГ с 2 компонентами в ближайшем контексте, в пределах 2-3 предложений или ее можно найти в заголовке, перечне ключевых слов и аннотации. Данное обстоятельство позволило предложить процедуру работы с ИГ в специализированном корпусе текстов, построенном в соответствии с предметной областью и конкретными интересами исследователя. Корпусные инструменты обеспечивают студентов/авторов L2 информацией, которая им необходима для того, чтобы идентифицировать и «распаковывать» эти конструкции в научных текстах своей предметной области, а также использовать их в своих собственных текстах.

Применение корпусных технологий дает возможность не только идентифицировать скрытую структуру ИГ и обнаружить реальные связи между ее элементами, но наблюдать ИГ в реальном контексте и в рамках автоматически найденных N-грамм.

Таким образом, мы рекомендовали ряд информационных технологий и сетевых ресурсов, потенциал которых может поддерживать и обеспечивать эффективность и корректность работы с важнейшим элементом научного текста.

СПИСОК ИСТОЧНИКОВ

1. Liu L., Li L. Noun Phrase Complexity in EFL Academic Writing: A Corpus-based Study of Postgraduate Academic Writing. // The Journal of Asia TEFL. 2016. Vol. 13. No. 1. Pp. 48–65. URL: <http://dx.doi.org/10.18823/asiatefl.2016.13.1.4.48> (дата обращения: 10.11.2021).



2. **Короткина И.Б.** Академическое письмо: процесс, продукт и практика. Москва: Юрайт, 2018. 295 с.
3. **Qian J., Krugly-Smolka E.** Chinese Graduate Students' Experiences with Writing a Literature Review // *TESL Canada Journal*. 2008. Vol. 26, No. 1. DOI: 10.18806/TESL.V26I1.391
4. **Yeh C.C.** New graduate students' perspectives on research writing in English: A case study in Taiwan // *Journal of Academic Language and Learning*, 2010. 4 (1), pp. A1–A12.
5. **Беляева Л.Н.** Английский язык в аспекте прикладного лингвистического исследования // Структурная и прикладная лингвистика. Вып. 7. СПб: Изд-во С.-Петерб. Ун-та, 2008. С. 159–174.
6. **Беляева Л.Н., Чернявская В.Е.** Technical writer in the framework of modern natural language processing tasks // *Журнал Сибирского федерального университета. Серия: Гуманитарные науки*. 2019. Т. 12, № 1. С. 20–31.
7. **Lauer M.** Corpus statistics meet the noun compound: Some empirical results // *Proc. of the 33rd Annual Meeting of the Association for Computational Linguistics*. Cambridge, Massachusetts, USA. 1995. Pp. 47–54.
8. **Guevara E.** A regression model of adjective-noun compositionality in distributional semantics // *Proc. of the ACL GEMS workshop*. 2010. Pp. 33–37.
9. **Gulordava K., Merlo P., Crabbe B.** Dependency length minimisation effects in short spans: a large-scale analysis of adjective placement in complex noun phrases // *Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics: Short Papers*. 2015. Pp. 477–482.
10. **Bos J., Nissim M.** Uncovering noun-noun compound relations by gamification // *Proc. of the 20th Nordic Conference of Computational Linguistics*. 2015. Pp. 251–255.
11. **Gellerstam M.** Translationese in Swedish novels translated from English // *Translation Studies in Scandinavia* / Wollin, L. and Lindquist, H. CWK Gleerup, Lund. 1986. No. 4. Pp. 88–95.
12. **Castilho S., Resende N., Mitkov R.** What Influences the Features of Features of Post-Editese? A Preliminary Study // *Proc. of the 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019)*, Temnikova I., Orasan C., Corpas Pastor G. and Mitkov R. (Eds.). Varna, Bulgaria, 2019, September 5 – 6. Pp. 19–27.
13. **Baker M.** Corpus linguistics and translation studies: Implications and applications // *Text and Technology: In Honour of John Sinclair* / Gill Francis and Elena Tognini-Bonelli (Eds.). John Benjamins Publishing Company, Netherlands, 1993. Pp. 233–252.
14. **Volansky V., Ordan N., Wintner Sh.** On the features of translationese // *Digital Scholarship in the Humanities*. 2013. No. 30 (1). Pp. 98–118. DOI: <https://doi.org/10.1093/lc/fqt031>
15. **Daems J., De Clercq O., Macken L.** Translationese and post-editese: How comparable is comparable quality? // *Linguistica Antverpiensia New Series – Themes in Translation Studies*. 2017. Vol. 16. Pp. 89–103.
16. **Toral A., Castilho S., Hu K., Way A.** Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation. 2018. URL: <https://arxiv.org/abs/1808.10432>. (дата обращения: 10.11.2021).
17. **Giacomini L.** Frame-based Lexicography: Presenting Multiword Terms in a Technical E-dictionary // *Proc. of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, Ljubljana: Ljubljana University Press. 2018. Pp. 309–318.
18. **Biber D., Conrad S., Leech J.** Longman grammar of spoken and written English. Harlow: Pearson Education Limited, 2007. 1204 p.
19. **Biber D., Gray B.** Grammatical change in the noun phrase: The influence of written language use // *English Language and Linguistics*. 2011. No. 15 (2). Pp. 223–250.
20. **Fernández-Domínguez J.** N+N compounding in English: Semantic categories and the weight of modifiers // *Brno Studies in English*. 2010. No. 36 (1). Pp. 47–76.
21. **Beliaeva L.** Applied Lexicography and Scientific Text Corpora // *Transactions on Business and Engineering Intelligent Applications*. Galina Setlak, Kassimir Markov (Ed.). Rzeszow, Poland: ITHEA, 2014. Pp. 55–63.
22. **Elliott Th.** Variation in use of nouns as nominal premodifiers in advanced student writing across academic disciplines. Graduate Theses and Dissertations. 2019. URL: <https://lib.dr.iastate.edu/etd/17445>. (дата обращения: 10.11.2021).
23. **Biber D., Gray B., Poonpon K.** Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? // *Tesol Quarterly*. 2011. No. 45 (1). Pp. 5–35.

24. **Parkinson J., Musgrave J.** Development of noun phrase complexity in the writing of English for Academic Purposes students // *Journal of English for Academic Purposes*. 2014. No. 14 (2). Pp. 48–59.
25. **Staples S., Egbert J., Biber D., Gray B.** Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre // *Written Communication*. 2016. No. 33(2). Pp. 149–183.
26. **Ansarifar A., Shahriari H., Pishghadam R.** Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics // *Journal of English for Academic Purposes*. 2018. No. 31. P. 58–71. DOI: 10.1016/j.jeap.2017.12.008
27. **Малаховская М.Л.** Субстантивные атрибутивные словосочетания в практике преподавания английского языка для академических целей // *Герценовские чтения. Иностранные языки: сборник научных трудов*. СПб.: Изд-во РГПУ им. А.И. Герцена, 2019. С. 310–313.
28. Peter Strutt's Delta Pages. "Noun Combinations." December 27, 2015. URL: <http://peterstrutt.co.uk/noun-combinations/> (дата обращения: 15.07.2021)
29. **Bauer L.** When is a sequence of two nouns a compound in English? // *English Language and Linguistics*. 1998. No. 2 (1). Pp. 65–86. DOI: 10.1017/S1360674300000691
30. **Linh N.M.** Noun-noun combinations in technical English. Master dissertation, English Language Studies Department, Suranaree University of Technology, 2010.
31. **Fries M.-H.** Teaching compound nouns in ESP: insights from cognitive semantics // *New developments in ESP teaching and learning research* / C. Sarré, S. Whyte (Eds). 2017. Pp. 93–107. DOI: <https://doi.org/10.14705/rpnet.2017.cssw2017.747>
32. **Adel A., Romer U.** Research on advanced student writing across disciplines and levels: Introducing the Michigan Corpus of Upper-level Student Papers // *International Journal of Corpus Linguistics*. 2012. No. 17 (1). Pp. 3–34.
33. **Anthony L.** Reflections and future directions in publishing research in English as an Additional Language: An Afterword // *Publishing Research in English as an Additional Language: Practices, Pathways and Potentials* / M. Cargill and S. Burgess (Eds.). Adelaide, Australia: University of Adelaide Press. 2017. Pp. 255–258.
34. **Anthony L.** Visualization in Corpus-Based Discourse Studies // *Corpus Approaches to Discourse: A Critical Review* / C. Taylor and A. Marchi (Eds.). Abingdon: UK. Routledge Press. 2018. Pp. 197–224.
35. **Anthony L.** Tools and strategies for Data-Driven Learning (DDL) // *Specialised English: New Directions in ESP and EAP Research and Practice* / K. Hyland, & L. Wong (Eds.). Abingdon: UK. Routledge Press. 2019. DOI: 10.4324/9780429492082-14
36. **Cargill M., Burgess S.** Introduction: Unpacking English for Research Publication Purposes [ER-PP] and the intersecting roles of those who research, teach and edit it // *Publishing Research in English as an Additional Language: Practices, Pathways and Potentials* / M. Cargill, S. Burgess (Eds.). The University of Adelaide Press. 2017. Pp. 1–12.
37. **Malakhovskaya M., Beliaeva L., Kamshilova O.** Teaching Noun-Phrase Composition in EAP/ESP Context: A Corpus-Assisted Approach to Overcome a Didactic Gap // *Journal of Teaching English for Specific and Academic Purposes*. 2021. Vol. 9, No 2. Pp. 257–266. DOI: <https://doi.org/10.22190/JTE-SAP2102257M>
38. **Беляева Л.Н.** Корпусные технологии извлечения терминологии в задачах переводной лексикографии // *Структурная и прикладная лингвистика*. Вып. 10. СПб.: Изд-во С.-Петербур. ун-та, 2014. С. 169–181.
39. **Беляева Л.Н.** Сетевые лингвистические технологии / Л.Н.Беляева, О.Н. Камшилова, Р.Г. Пиотровская. СПб: Изд-во РГПУ им. А.И. Герцена, 2019. 111 с.
40. **Kunanets N., Levchenko O., Hadzalo A.** The Application of AntConc Concordancer in Linguistic Researches // *Proc. of 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv, 2018. Pp. 144–147. DOI: 10.1109/STC-CSIT.2018.8526591

REFERENCES

[1] **L. Liu, L. Li**, Noun Phrase Complexity in EFL Academic Writing: A Corpus-based Study of Postgraduate Academic Writing, *The Journal of Asia TEFL*, 13 (1) (2016) 48–65. URL: <http://dx.doi.org/10.18823/asiatefl.2016.13.1.4.48> (accessed 10.11.2021).



- [2] **I.B. Korotkina**, *Akademicheskoye pismo: protsess, produkt i praktika* [Academic writing: process, product and practice]. Moscow: Yurayt, 2018. 295 p.
- [3] **J. Qian, E. Krugly-Smolksa**, Chinese Graduate Students' Experiences with Writing a Literature Review, *TESL Canada Journal*, 26 (1) (2008). DOI: 10.18806/TESL.V26I1.391
- [4] **C.C. Yeh**, New graduate students' perspectives on research writing in English: A case study in Taiwan, *Journal of Academic Language and Learning*, 4 (1) (2010) A1–A12.
- [5] **L.N. Belyayeva**, *Angliyskiy yazyk v aspekte prikladnogo lingvisticheskogo issledovaniya* [English in the aspect of applied linguistic research], *Strukturnaya i prikladnaya lingvistika* [Structural and Applied Linguistics], Vol. 7. SPb: S.-Peterb. University press, 2008. Pp. 159–174.
- [6] **L.N. Belyayeva, V.Ye. Chernyavskaya**, Technical writer in the framework of modern natural language processing tasks, *Zhurnal Sibirskogo federalnogo universiteta. Seriya: Gumanitarnyye nauki* [Journal of Siberia Federal University: Humanities], 12 (1) (2019) 20–31.
- [7] **M. Lauer**, Corpus statistics meet the noun compound: Some empirical results, *Proc. of the 33rd Annual Meeting of the Association for Computational Linguistics*. Cambridge, Massachusetts, USA. 1995. Pp. 47–54.
- [8] **E. Guevara**, A regression model of adjective-noun compositionality in distributional semantics, *Proc. of the ACL GEMS workshop*. 2010. Pp. 33–37.
- [9] **K. Gulordava, P. Merlo, B. Crabbe**, Dependency length minimisation effects in short spans: a large-scale analysis of adjective placement in complex noun phrases, *Proc. of the 53rd Annual Meeting of the Association for Computational Linguistics: Short Papers*. 2015. Pp. 477–482.
- [10] **J. Bos, M. Nissim**, Uncovering noun-noun compound relations by gamification, *Proc. of the 20th Nordic Conference of Computational Linguistics*. 2015. Pp. 251–255.
- [11] **M. Gellerstam**, Translationese in Swedish novels translated from English, *Translation Studies in Scandinavia / Wollin, L. and Lindquist, H. CWK Gleerup, Lund*. 4 (1986) 88–95.
- [12] **S. Castilho, N. Resende, R. Mitkov**, What Influences the Features of Features of Post-Editese? A Preliminary Study, *Proc. of the 2nd Workshop on Human-Informed Translation and Interpreting Technology (HiT-IT 2019)*, Temnikova I, Orasan C., Corpas Pastor G. and Mitkov, R. (Eds.). Varna, Bulgaria, 2019, September 5 – 6. Pp. 19–27.
- [13] **M. Baker**, Corpus linguistics and translation studies: Implications and applications, *Text and Technology: In Honour of John Sinclair / Gill Francis and Elena Tognini-Bonelli* (Eds). John Benjamins Publishing Company, Netherlands, 1993. Pp. 233–252.
- [14] **V. Volansky, N. Ordan, Sh. Wintner**, On the features of translationese, *Digital Scholarship in the Humanities*, 30 (1) (2013) 98–118. DOI: <https://doi.org/10.1093/lc/fqt031>
- [15] **J. Daems, O. De Clercq, L. Macken**, Translationese and post-editese: How comparable is comparable quality?, *Linguistica Antverpiensia New Series – Themes in Translation Studies*, 16 (2017) 89–103.
- [16] **A. Toral, S. Castilho, K. Hu, A. Way**, Attaining the Unattainable? Reassessing Claims of Human Parity in Neural Machine Translation. 2018. Available at: <https://arxiv.org/abs/1808.10432> (accessed 10.11.2021).
- [17] **L. Giacomini**, Frame-based Lexicography: Presenting Multiword Terms in a Technical E-dictionary, *Proc. of the XVIII EURALEX International Congress: Lexicography in Global Contexts*, Ljubljana: Ljubljana University Press. 2018. Pp. 309–318.
- [18] **D. Biber, S. Conrad, J. Leech**, *Longman grammar of spoken and written English*. Harlow: Pearson Education Limited, 2007. 1204 p.
- [19] **D. Biber, B. Gray**, Grammatical change in the noun phrase: The influence of written language use, *English Language and Linguistics*, 15 (2) (2011) 223–250.
- [20] **J. Fernández-Domínguez**, N+N compounding in English: Semantic categories and the weight of modifiers, *Brno Studies in English*, 36 (1) (2010) 47–76.
- [21] **L. Beliaeva**, *Applied Lexicography and Scientific Text Corpora*, *Transactions on Business and Engineering Intelligent Applications*. Galina Setlak, Kassimir Markov (Ed.). Rzeszow, Poland: ITHEA, 2014. Pp. 55–63.
- [22] **Th. Elliott**, Variation in use of nouns as nominal premodifiers in advanced student writing across academic disciplines. *Graduate Theses and Dissertations*. 2019. Available at: <https://lib.dr.iastate.edu/etd/17445> (accessed 10.11.2021).
- [23] **D. Biber, B. Gray, K. Poonpon**, Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?, *Tesol Quarterly*, 45 (1) (2011) 5–35.

- [24] **J. Parkinson, J. Musgrave**, Development of noun phrase complexity in the writing of English for Academic Purposes students, *Journal of English for Academic Purposes*, 14 (2) (2014) 48–59.
- [25] **S. Staples, J. Egbert, D. Biber, B. Gray**, Academic writing development at the university level: Phrasal and clausal complexity across level of study, discipline, and genre, *Written Communication*, 33 (2) (2016) 149–183.
- [26] **A. Ansarifar, H. Shahriari, R. Pishghadam**, Phrasal complexity in academic writing: A comparison of abstracts written by graduate students and expert writers in applied linguistics, *Journal of English for Academic Purposes*, 31 (2018) 58–71. DOI: 10.1016/j.jeap.2017.12.008
- [27] **M.L. Malakhovskaya**, Substantivnyye atributivnyye slovosochetaniya v praktike prepodavaniya angliyskogo yazyka dlya akademicheskikh tseley [Attributive noun groups in teaching English for academic purposes], *Gertsenovskiy chteniya. Inostrannyye yazyki: sbornik nauchnykh trudov* [Proc. of Herzen Readings. Foreign Languages]. SPb.: Herzen University Press, 2019. Pp. 310–313.
- [28] Peter Strutt's Delta Pages. "Noun Combinations." December 27, 2015. Available at: <http://peterstrutt.co.uk/noun-combinations/> (accessed 15.07.2021).
- [29] **L. Bauer**, When is a sequence of two nouns a compound in English?, *English Language and Linguistics*, 2 (1) (1998) 65–86. DOI: 10.1017/S1360674300000691
- [30] **N.M. Linh**, Noun-noun combinations in technical English. Master dissertation, English Language Studies Department, Suranaree University of Technology, 2010.
- [31] **M.-H. Fries**, Teaching compound nouns in ESP: insights from cognitive semantics, *New developments in ESP teaching and learning research* / C. Sarré, S. Whyte (Eds). 2017. Pp. 93–107. DOI: <https://doi.org/10.14705/rpnet.2017.cssw2017.747>
- [32] **A. Adel, U. Romer**, Research on advanced student writing across disciplines and levels: Introducing the Michigan Corpus of Upper-level Student Papers, *International Journal of Corpus Linguistics*, 17 (1) (2012) 3–34.
- [33] **L. Anthony**, Reflections and future directions in publishing research in English as an Additional Language: An Afterword, *Publishing Research in English as an Additional Language: Practices, Pathways and Potentials* / M. Cargill and S. Burgess (Eds.). Adelaide, Australia: University of Adelaide Press. 2017. Pp. 255–258.
- [34] **L. Anthony**, Visualization in Corpus-Based Discourse Studies, *Corpus Approaches to Discourse: A Critical Review* / C. Taylor and A. Marchi (Eds.). Abingdon: UK. Routledge Press. 2018. Pp. 197–224.
- [35] **L. Anthony**, Tools and strategies for Data-Driven Learning (DDL), *Specialised English: New Directions in ESP and EAP Research and Practice* / K. Hyland, & L. Wong (Eds.). Abingdon: UK. Routledge Press. 2019. DOI: 10.4324/9780429492082-14
- [36] **M. Cargill, S. Burgess**, Introduction: Unpacking English for Research Publication Purposes [ER-PP] and the intersecting roles of those who research, teach and edit it, *Publishing Research in English as an Additional Language: Practices, Pathways and Potentials* / M. Cargill, S. Burgess (Eds.). The University of Adelaide Press. 2017. Pp. 1–12.
- [37] **M. Malakhovskaya, L. Beliaeva, O. Kamshilova**, Teaching Noun-Phrase Composition in EAP/ESP Context: A Corpus-Assisted Approach to Overcome a Didactic Gap, *Journal of Teaching English for Specific and Academic Purposes*, 9 (2) (2021) 257–266. DOI: <https://doi.org/10.22190/JTE-SAP2102257M>
- [38] **L.N. Belyayeva**, Korpusnyye tekhnologii izvlecheniya terminologii v zadachakh perevodnoy leksikografii [Corpus technologies of term extraction for lexicography tasks], *Strukturnaya i prikladnaya lingvistika* [Structural and Applied Linguistics], Vol. 10. SPb: SPb. University Press, 2014. Pp. 169–181.
- [39] **L.N. Belyayeva**, Setevyye lingvisticheskiye tekhnologii [Web linguistic technologies] / L.N. Belyayeva, O.N. Kamshilova, R.G. Piotrovskaya. SPb: Herzen University Press, 2019. 111 p.
- [40] **N. Kunanets, O. Levchenko, A. Hadzalo**, The Application of AntConc Concordancer in Linguistic Researches, *Proc. of 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv, 2018. Pp. 144–147. DOI: 10.1109/STC-CSIT.2018.8526591

СВЕДЕНИЯ ОБ АВТОРАХ / THE AUTHORS

Беляева Лариса Николаевна

Beliaeva Larisa N.

E-mail: lauranbel@gmail.com

ORCID: <https://orcid.org/0000-0002-8622-4595>

Камшилова Ольга Николаевна

Kamshilova Olga N.

E-mail: onkamshilova@gmail.com

ORCID: <https://orcid.org/0000-0002-1488-2206>

Малаховская Мария Львовна

Malakhovskaya Maria L.

E-mail: lmalakh2001@mail.ru

ORCID: <https://orcid.org/0000-0001-7590-7182>

Статья поступила в редакцию 18.11.2021; одобрена после рецензирования 24.12.2021; принята к публикации 26.12.2021.

The article was submitted 18.11.2021; approved after reviewing 24.12.2021; accepted for publication 26.12.2021.