

Выявление номинализованных конструкций в параллельных текстах патентных документов на русском и немецком языках

Detection of nominalized structures in parallel patent texts in Russian and in German

Кожунова О. С. (okozhunova@ipiran.ru)

Институт проблем информатики Российской академии наук

Исследуется явление номинализации в двуязычной ситуации (русский-немецкий) с привлечением результатов сопоставительных исследований для трех языков: русский — английский — немецкий, а также способ идентификации параллельности текстов для патентной сферы и типы трансформаций.

1. Введение

На современном этапе создания систем обработки естественного языка для информационных технологий особое значение приобретает разработка методик анализа параллельных текстов на нескольких языках. При этом возникает множество задач, связанных с их адекватной интерпретацией и применением, в первую очередь, задачи машинного перевода и обработки знаний [1]. Проблема извлечения и обработки знаний открывает перспективы развития интеллектуальных направлений компьютерной лингвистики, поскольку ее основной акцент смещен в сторону глубинных представлений языка, в которых используются как грамматические (морфологические и синтаксические), так и семантические атрибуты для описания языковых объектов. В данной работе исследования параллельных текстов рассматриваются именно в ракурсе этой проблемы.

Как показано в [2] ключевой задачей при разработке методов сопоставления параллельных текстов является выявление и детальное описание тех языковых трансформаций, которые имеют место при переводе естественно-языковых конструкций с одного языка на другой, поскольку далеко не всегда некоторое содержание передается структурно-подобными средствами в текстах на разных языках. При этом особое внимание должно уделяться выявлению признаков «глагольности» и «номинативности» [2] в сопоставляемых языковых структурах параллельных текстов, поскольку это важно для последующей разработки алгоритмов анализа текстов с использованием методов машинного обучения. Сравнительное исследование употребления различных частей речи в параллельных текстах на раз-

ных языках дает основу для выявления и описания языковых трансформаций, при этом центральной трансформацией является *номинализация*. Явление номинализации было исследовано в ряде работ отечественных и зарубежных лингвистов [3, 4, 5, 6]. Ближе всего к нашему пониманию этого явления определения, данные в работе [4]: «конструкции... называются номинализованными в том смысле, что их естественно рассматривать как результат номинализации конструкций с предикативным употреблением глаголов и прилагательных» и [5]: «номинализация — это синтаксический процесс, который соотносит предложения с именными группами». Основная цель наших исследований выявление номинализованных конструкций в патентных текстах на русском и немецком языках, прежде всего, в формулах изобретений и сопоставительное описание глагольно-именных межъязыковых трансформаций. Центральная задача наших исследований — выявление тех ситуаций, когда фраза с личным глаголом в тексте на одном языке превращается в именную фразу в тексте на другом языке и обратно.

2. Определение эквивалентности языковых структур при сопоставлении патентных текстов и выявление номинализации

Тексты патентных изобретений, по существу, представляют собой особый жанр, поскольку к ним предъявляются строго регламентированные требования как на содержательном уровне (тематика, сте-

пень новизны и актуальности текста определенной предметной области, иллюстративность, практическое значение и т.п.), так и в вопросе оформления, к соблюдению определенной структуры, прозрачности и четкости изложения и т. д. Так, например, необходимым условием патентного описания является четкое перечисление характеристик изобретения по нумерованным пунктам. Казалось бы, это условие должно облегчать их анализ при сопоставлении с аналогичным описанием на другом языке. Однако жесткость структуры патентных формул влечет за собой осложнение процедуры их сравнения, так как нумерация пунктов часто не совпадает и условие параллельности текстов не соблюдается.

Поэтому при анализе формул изобретения патентных текстов на русском, немецком и английском языках на основании отобранного массива текстов патентов 2006 года (455 текстовых представлений) была выявлена необходимость в разра-

ботке подхода к определению параллельности таких текстовых описаний и предварительной типологии анализируемых текстов.

Для первичного отбора параллельных текстов на трех языках был использован инструмент аккумуляции и индексации патентов MIMOSA V5, который представляет собой базу данных кратких патентных описаний. Внутри каждого описания размещены ссылки на внешние ресурсы, содержащие формулы изобретения и соответствующие тексты патентов. С целью проведения синтаксического разбора патентных формул изобретения на русском и английском языках применялся программный пакет Dwarf 2.0. Этот синтаксический анализатор позволяет проводить грамматический разбор текстов и строить деревья разметки с выделением соответствующих атрибутов. Далее в тексте приведены примеры, в которых используются результаты работы Dwarf 2.0 (пример 1).

(1) *Формула изобретения на немецком языке:*

1. **Vorrichtung zum Sterilisieren von Getränkebehältern durch Strömungsbeaufschlagung mit einem Gemisch aus Luft und H2O2-Dampf, mit einem H2O2-Verdampfer, der einen luftdurchströmten Verdampfungsraum (4) mit beheizten Wänden aufweist*, gegen die H2O2 von einer Düseneinrichtung (7,9) gesprüht wird, dadurch gekennzeichnet, dass die Düseneinrichtung eine engen Luftstrahl (8) einblasende Luftdüse (7) und eine H2O2 Düse (9) aufweist, die zur Erzeugung eines H2O2-Strahles (10) im Abstand zum Luftstrahl und auf diesen gerichtet ausgebildet ist.*

Фрагмент синтаксического разбора¹

1	*Top*	*Top*	1	_	/
2	vorrichtung	vorrichtung	1	vp	/nn_verb, sg, nom
3	zum	zum	4	prep	/prep_dativ
4	sterilisieren	sterilisieren	6	vp	/nn_verb, sg, dat
5	von	von	6	prep	/prep_dativ
6	getraenkebehaeltern	getraenkebehaelter	8	np	/nn, pl, dat
7	durch	durch	8	prep	/prep_accus
8	Stromungsmungsbeaufschlagung	Stromungsbeaufschlagung	11	np	/nn_verb, sg, acc
9	mit	mit	10	prep	/prep_dativ
10	einem	einem	11	det	/det, indef, sg, dat
11	gemisch	gemisch	13	np	/nn, sg, acc
12	aus	aus	13	prep	/prep_dativ
13	luft	luft	11	np	/nn/sg/trd/heu/
14	und	und	15	cnj	/ cnj
15	h202-dampf	h202-dampf	18	np	/nn_comp, sg, acc
16	mit	mit	17	prep	/ prep_dativ
17	einem	einem	18	det	/det, indef, sg, dat
18	h202-verdampfer	h202-verdampfer	21	np	/nn_comp, sg, acc
19	der	der	20	det	/det, def, sg, acc
20	einen	einen	21	det	/det, indef, sg, acc
21	luftdurchstroemten	luftdurchstromte	22	adj	/ adj, sg, acc
22	verdampfungsraum	verdampfungsraum	21	np	/nn, sg, acc
23	mit	mit	24	prep	/ prep_dativ, управл датив
24	beheizten	beheizt	25	adj	/ adj, pl, dativ
25	waenden	wand	24	np	/nn/sg/trd/heu/
26	aufweist	aufweisen	11	vp	/vb, sg, 3p

¹ Разбор выполнен вручную с использованием грамматических атрибутов и средств разметки Dwarf 2.0: <http://cs.isa.ru:10000/troll/>

В ходе эксперимента по анализу немецких, русских и английских патентных текстов было выявлено, что формулы изобретения существенным образом трансформируются структурно, что также отражается на порядке описания содержательной компоненты.

Например, отдельные фрагменты формулы изобретения на одном из рассматриваемых языков расширяются, включают в себя более детальные описания, или, напротив, сужаются, часто происходит удаление части контента или его дополнение:

В параллельных текстах примера 1 можно проследить следующие русско-немецкие лексико-семантические и синтаксические трансформации:

(a) *Getränkebehältern* -> емкостей для напитков
N [dat, masc, pl] -> N [acc, fem, pl] + Prep + N [gen, masc, pl]

(b) *durch Strömungsbeaufschlagung* -> путем воздействия
Prep + N [acc, fem, sg] -> Prep + N [gen, neut, sg]

(c) *luftdurchströmten* -> оmyваемую воздухом
Comp_part [acc, masc, sg] -> Part [acc, fem, sg] + N [instr, masc, sg]

(d) *Verdampfungsraum* -> испарительную камеру
N [acc, masc, sg] -> Adj [acc, fem, sg] + N [acc, fem, sg]

Формула изобретения на русском языке:

1. *Устройство для стерилизации емкостей для напитков путем воздействия на них потоком смеси воздуха и пара перекиси водорода, содержащее испаритель H₂O₂, имеющий оmyваемую воздухом испарительную камеру (4) с обогреваемыми стенками, * которые с помощью соплового устройства (7, 9) обрызгиваются перекисью водорода, отличающиеся тем, что сопловое устройство содержит только одно воздушное сопло (7), выполненное с возможностью вдувания тангенциально в камеру узкой воздушной струи (8), и только одно сопло (9) для перекиси водорода, выполненное с возможностью формирования сплошной струи жидкости, имеющей диаметр примерно в пределах равных нескольким десяткам миллиметра, причем сопло для перекиси водорода установлено в корпусном блоке (3) с возможностью формирования струи жидкости (10) в поперечном к воздушной струе (8) направлении и на расстоянии от нее.

Фрагмент синтаксического разбора^{2}:

1	*Тор*	*Тор*	1	_	/
2	устройство	устройство	1	np	/nn/sg/neu/nom/
3	для	для	4	prep	/prp/
4	стерилизации	стерилизация	2	prepn	/nn/sg/fem/gen/
5	емкостей	емкость	4	gen	/nn/pl/fem/gen/
6	для	для	7	prep	/prp/
7	напитков	напиток	5	prepn	/nn/pl/msc/gen/
8	путем	путь	9	prep	/nn/sg/msc/ins/
9	воздействия	воздействие	7	prepn	/nn/sg/neu/gen/
10	на	они	11	prep	/prp/
11	них	они	9	prepn	/prn/pl/acc/trd/
12	потоком	поток	1	np	/nn/sg/msc/ins/
13	смеси	смесь	12	gen	/nn/sg/fem/gen/
14	воздуха	воздух	13	gen	/nn/sg/msc/gen/
15	и	и	14	conj	/cnj/
16	пара	пар	14	homo	/nn/sg/msc/gen/
17	перекиси	перекись	16	gen	/nn/sg/fem/gen/
18	водорода	водород	17	gen	/nn/sg/msc/gen/
20	содержащее	содержать	1	adj	/ptp/sg/neu/nom/prs/act/
21	испаритель	испаритель	20	acc	/nn/sg/msc/nom/
22	н ₂ о ₂	н ₂ о ₂	1	misc	/
25	имеющий	иметь	23	adj	/ptp/sg/msc/nom/prs/act/
26	оmyваемую	оmyвать	29	ptp	/ptp/sg/fem/acc/prs/psv/
27	воздухом	воздух	26	ins	/nn/sg/msc/ins/
28	испарительную	испарительный	29	adj	/adj/sg/fem/acc/
29	камеру	камера	25	acc	/nn/sg/fem/acc/
33	с	с	35	prep	/prp/
34	обогреваемыми	обогревать	35	adj	/ptp/pl/ins/prs/psv/
35	стенками	стенка	23	prepn	/nn/pl/fem/ins/

² Разбор выполнен при помощи инструмента для синтаксического разбора Dwarf 2.0: <http://cs.isa.ru:10000/troll/>

Трансформации типов (a)–(d) назовем *предложно-номинативными*. Они встречаются в патентных описаниях формул изобретений среди русско-немецких двоек параллельных текстов с доминирующей частотой (~60 %). Таким образом, преобразование составных отглагольных существительных (немецкий язык) во фразовые номинативные структуры (русский язык) в формулах изобретения патентных параллельных текстов является наиболее распространенным. При чем подобные трансформации происходят не только на уровне семантики различных языковых структур, но и на лексическом уровне (как это видно из структурных формул приведенных выше), поскольку, как правило, одна языковая единица немецкого языка преобразуется в несколько языковых единиц русского языка.

(e) *mit beheizten Wänden* -> *с обогреваемыми стенками*

Prep + Part [dat, fem, pl] + N [dat, fem, pl] -> Prep + Part [instr, fem, pl] + N [[instr, fem, pl]

Менее частотные трансформации типа (e) назовем *адъективно-номинативными*. Причем встречаемость таких трансформаций, как показал анализ экспериментального массива формул изобретений, как правило, зависит от исходной предметной области описываемых изобретений. Например, в описаниях патентов в области химии и медицины доминируют именно *адъективно-номинативные* трансформации:

die lösliche Guanylatcyclase -> *растворимой гуанилатциклазы*

verzweigtes Halogenalkoxy -> *разветвленный галогеналкил*

(f) *der einen [a] aufweist (который имеет) -> имеющий [a]*

Pro [rel, masc, sg] + Art [indef, masc, acc] + [a] + Vb [3 ps, sg, pres] -> Part [nom, masc, sg] + [a]

Что касается трансформаций типа (f) (назовем их *структурно-вербальными*), то они носят другой функционально-семантический характер. То, что было относительным местоимением и личной формой глагола в немецком языке становится причастием в русском языке. Таким образом, в структурно-вербальных трансформациях осуществляются не только лексико-семантические преобразования, но и синтаксические.

(2) **Формула изобретения на немецком языке:**

2. Vorrichtung nach Anspruch 1, dadurch gekennzeichnet, dass der Verdampfungsraum (4) mit dem Auslass(1) des Verdampfers über beheizte Kanäle (5) verbunden ist.

Формула изобретения на русском языке:

3. Устройство по п.1, отличающееся тем, что испарительная камера (4) соединена с выходным отверстием (1) испарителя обогреваемыми каналами (5).

В примере 2 также присутствуют выделенные в примере 1 трансформации. Тем не менее, остановимся на наиболее интересном преобразовании:

[a] соединена [b] -> [a] [b] verbunden ist

[a] + *Part [short, nom, fem, sg] + [b] -> [a] + [b] + Part [II, sg, masc] + Vb [3 ps, sg, pres]*

Аналогичная трансформация была рассмотрена в примере 1, случай (f). Здесь также присутствует не только лексико-семантическое преобразование, но и синтаксическое: краткое причастие на русском языке трансформируется в составную глагольную структуру на немецком языке.

Большое количество формул изобретения аналогичных приведенным примерам позволяет сделать вывод о том, что их сопоставительный анализ следует проводить на основе выделения содержательных объектов. То есть при первичном отборе параллельных текстов и определении их параллельности необходимо опираться не на синтаксическую структуру, а на выделенные семантические объекты [7, 8]. Поэтому на стадии предварительного анализа текстов мы подразделяем их на собственно параллельные, непараллельные и *концептуально-сопоставимые*³. Под *параллельными текстами* в данном случае понимаются тексты, которые являются эквивалентными и адекватными переводам с одного языка на другой, то есть совпадающими по объему фактической информации в них представленной. Тексты, в которых описывается один и тот же контент, но которые не являются эквивалентными переводами и не совпадают по объему, мы называем *концептуально-сопоставимыми текстами*.

В зависимости от того, с какими текстами мы имеем дело, варьируется и методика их анализа. В частности, при сопоставлении параллельных текстов наиболее интересен последовательный контрастивный анализ предложений с целью обнаружения трансформаций, которые, как правило, регулярные и предсказуемые:

(3)

- **Формула изобретения на немецком языке:**
Verfahren zur Epoxidierung einer organischen Verbindung mit wenigstens einer C C-Doppelbindung mit Wasserstoffperoxid in Gegenwart wenigstens einer katalytisch aktiven Verbindung und wenigstens eines Lösungsmittels, dadurch gekennzeichnet

³ Термин *концептуально-сопоставимые* введен в данной работе впервые и является результатом наших исследований структурных особенностей параллельных текстов формул изобретения в патентных документах.

net, dass ein Produktgemisch umfassend a-Hydroperoxyalkohole unter Einsatz wenigstens eines Reduktionsmittels reduziert wird.

- **Формула изобретения на русском языке:** Способ эпоксидирования пропена путем взаимодействия с перекисью водорода в присутствии, по меньшей мере, одного каталитически активного соединения, включающего цеолитный катализатор, и, по меньшей мере, одного органического растворителя с использованием разделения продуктов, отличающийся тем, что смесь продукта, содержащую а-гидропероксипропанола, восстанавливают с применением, по меньшей мере, одного восстановителя, причем а-гидропероксипропанола восстанавливают до соответствующего пропиленгликоля.
- **Формула изобретения на английском языке:** A process for the epoxidation of an organic compound having at least one C-C double bond by means of hydrogen peroxide in the presence of at least one catalytically active compound and at least one solvent, wherein a product mixture comprising [alpha]-hydroperoxyalcohols is reduced using at least one reducing agent.

(a) Verfahren zur Epoxidierung → Способ эпоксидирования → A process for the epoxidation
N [verb, nom, neutr, sg] + Prep [zu+der, dat, comp, fem, sg] + N [dat, fem, sg] → N [nom, masc, sg] + N [gen, neutr, sg] → Art [indef, sg] + N [com, sg] + Prep + Art [def, 0] + N [com, sg]

(b) ein Produktgemisch → смесь продукта → a product mixture
Art [indef, masc, nom, sg] + N [comp, nom, neutr, sg] → N [nom, fem, sg] + N [gen, masc, sg] → Art [indef, sg] + N [com, sg] + N [com, sg]

(c) dadurch gekennzeichnet → отличающийся тем → wherein
Pron + Part [II f, masc, sg] → Part [nom, masc, sg] + Pron [instr, masc, sg] → Adv

В примере 3 приведены трансформации, которые обнаружены в тройке параллельных текстов. Для троек параллельных текстов патентных формул изобретений наиболее частотными типами трансформаций оказались преобразования (a) и (b), то есть предложно-номинативные и адъективно-номинативные (порядка 55 %), реже встречаются трансформации типа (c), которые мы назовем предложно-адъективными (около 30 %). Как уже было отмечено в комментариях к примерам 1 и 2, предложно-номинативные и адъективно-номинативные трансформации являются наиболее частотными для русско-немецких параллельных текстов. В данном примере такие трансформации наиболее характерны для параллельных троек тек-

стов, что объясняется сходством грамматических структур в немецком и английском языках.

Ниже приведен пример не параллельных, а сопоставимых текстов на немецком и русском языках. В таких текстах уже сложнее выявлять трансформации, следуя жестко регламентированной структуре патентных описаний (нумерация пунктов не совпадает, перевод текста может содержать более детальную или, напротив, более сжатую информацию, ключевые понятия могут быть заменены на синонимичные оригинальным терминам и т.д.). В частности, в примере 4 пункт б формулы изобретения на русском языке соответствует пункту 5 формулы изобретения на немецком языке, причем немецкий вариант более сжат по сравнению со своим русскоязычным аналогом.

(4) Формула изобретения на русском языке:

4. Способ по одному из пп.1-3, отличающийся тем, что измеренное отклонение передают на управляющее устройство, которое рассчитывает необходимое изменение расстояния и инициирует изменение расстояния.

5. Способ по п.1, отличающийся тем, что экструдуют пластмассовые трубы.

6. Способ изготовления экструдированных изделий путем экструдирования расплавленного жгута из экструдера, формирования расплавленного жгута в инструменте в расплавленный шланг, охлаждения и калибрования экструдированного изделия и его перемещения с помощью тянущего гусеничного устройства, содержащего по меньшей мере две гусеничные ленты, а также резки экструдированного изделия на отдельные части, отличающийся тем, что расстояние между гусеничными лентами и продольной осью гусеничного тянущего устройства в соответствии с данными измеренного изменения диаметра экструдированного изделия автоматически устанавливают в зависимости от изменения таким образом, чтобы ось симметрии экструдированного изделия совпала с продольной осью гусеничного тянущего устройства.

Формула изобретения на немецком языке:

4. Verfahren nach einem der Ansprüche 1 bis 3, dadurch gekennzeichnet, dass die gemessene Auslenkung einer Steuerung zugeführt wird, die die notwendige Abstandsänderung errechnet und die Abstandsänderung veranlasst.

5. Verfahren zum Herstellen von Extrusionsprodukten, insbesondere von Kunststoffrohren, mittels Extrudieren eines Schmelzestrangs aus einem Extruder, Formen des Schmelzestrangs in einem Werkzeug, insbesondere in einem Rohrwerkzeug zu einem Schmelzschlauch, Abkühlen und Kalibrieren des Extrusionsprodukts und Abziehen desselben mit einem Raupenabzug mit mindestens zwei Raupenkettens, sowie Trennen des Extrusionsprodukts in einzelne Stücke, dadurch gekennzeichnet, dass der Abstand der Raupenkettens von einer Längsachse des Raupenabzugs nach Maßgabe einer gemessenen Durchmesseränderung des Extrusionsprodukts in Abhängigkeit von dieser automatisch so eingestellt wird, dass die Symmetrieachse des Extrusionsprodukts mit der Längsachse des Raupenabzugs übereinstimmt.

Поэтому если параллельные тексты имеют смысл сопоставлять и анализировать последовательно, по предложениям, то концептуально-сопоставимые тексты анализируются с опорой на содержательные объекты, то есть концепты, релевантные для данной предметной области (сами концепты, их связи, характеристики и т.п.) [7]. Такой подход оправдывает себя уже потому, что в первой сотне патентных описаний (из 455 за один год) на трех языках

большинство составили именно концептуально-сопоставимые тексты (порядка 70 %).

В программе индексации и поиска патентов (MI-MOSA V5) такие тексты отмечены как эквивалентные, однако, при этом из-за различной длины и содержания текстов существенно затруднен поиск, анализ и сравнение значимых концептов. Поэтому при таком сопоставлении формул изобретения патентов на разных языках предлагается использовать предварительно сформированное лексико-семантическое представление, в котором содержатся соответствующие концепты со всеми возможными вербальными реализациями, синонимическим рядом и релевантными связями с другими концептами [8].

В примере 5 приведены фрагменты концептуально-сопоставимых текстов на немецком и русском языках:

(5) Формула изобретения на немецком языке:

5. Verfahren zum Herstellen von Extrusionsprodukten, insbesondere von Kunststoffrohren, mittels Extrudieren eines Schmelzestrangs aus einem Extruder, Formen des Schmelzestrangs in einem Werkzeug, insbesondere in einem Rohrwerkzeug zu einem Schmelzschlauch, Abkühlen und Kalibrieren des Extrusionsprodukts und Abziehen desselben mit einem Raupenabzug mit mindestens zwei Raupenkettens, sowie Trennen des Extrusionsprodukts in einzelne Stücke, dadurch gekennzeichnet, dass der Abstand der Raupenkettens von einer Längsachse des Raupenabzugs nach Massgabe einer gemessenen Durchmesseränderung des Extrusionsprodukts in Abhängigkeit von dieser automatisch so eingestellt wird, dass die Symmetrieachse des Extrusionsprodukts mit der Längsachse des Raupenabzugs übereinstimmt.

6. Raupenabzug zum Abziehen von Extrusionsprodukten, insbesondere von Kunststoffrohren (8), mit mindestens zwei Raupenkettens (10,10'), wobei die Raupenkettens (10,10') mindestens während des Abziehens symmetrisch zu einer Längsachse (12) des Raupenabzugs (6) anordenbar und an das Extrusionsprodukt andrückbar sind, sowie mit Mitteln zum Ändern des Abstands zwischen einer Anzahl von Raupenkettens (10,10') und der Längsachse (12) des Raupenabzugs (6), dadurch gekennzeichnet, dass Mittel vorgesehen sind, um Auslenkungen von Raupenkettens (10, 10') orthogonal zur Längsachse (12) zu messen

Формула изобретения на русском языке:

5. Способ по п. 1, отличающийся тем, что экструдируют пластмассовые трубы.

6. Способ изготовления экструдированных изделий путем экструдирования расплавленного жгута из экструдера, формования расплавленного жгута в инструменте в расплавленный шланг, охлаждения и калибрования экструдированного изделия и его перемещения с помощью тянущего гусеничного

устройства, содержащего по меньшей мере две гусеничные ленты, а также резки экструдированного изделия на отдельные части, отличающийся тем, что расстояние между гусеничными лентами и продольной осью гусеничного тянущего устройства в соответствии с данными измеренного изменения диаметра экструдированного изделия автоматически устанавливаются в зависимости от изменения таким образом, чтобы ось симметрии экструдированного изделия совпала с продольной осью гусеничного тянущего устройства.

В примере 5 явно прослеживается несовпадение контента одинаково пронумерованных пунктов при практически полном соответствии пункта 5 немецкоязычного варианта и пункта 6 русскоязычного варианта формулы изобретения. В русском варианте доминируют назывные предложения с распространениями, в то время как в немецком тексте в основном представлены глагольные структуры:

- Verfahren zum Herstellen von Extrusionsprodukten, insbesondere von Kunststoffrohren,.....
- Способ изготовления экструдированных изделий путем экструдирования расплавленного жгута из экструдера.....

Таким образом, подобные глагольно-именные соответствия, выраженные синтаксическими структурами на разных языках, можно назвать функционально-синонимическими [9, 10].

3. Заключение

В работе исследуется вопрос выявления номинализации в параллельных и концептуально-сопоставимых текстах формул изобретения патентов на трех языках (русском, английском и немецком), а также способ идентификации параллельности текстов для патентной сферы и типы трансформаций для русско-немецких и русско-немецко-английских текстовых представлений. Кроме того, анализируется структура патентных представлений с точки зрения извлечения лингвистической информации, а именно, концепты различных предметных областей, их характеристики, связи с другими концептами, языковые структуры, в которые они включены, трансформации, в которых они участвуют чаще всего и т. д.

Было выявлено, что наиболее частотными русско-немецкими трансформациями являются глагольно-именные трансформации. В различных типах текстов (параллельных и концептуально-сопоставимых, двойках и тройках параллельных текстов) доминируют определенные типы преобразований. В частности, наиболее распространенным является преобразование составных отглагольных существительных (немецкий язык) во фразовые

номинативные структуры (русский язык). Кроме того, было обнаружено, что в русско-немецких параллельных текстах осуществляются не только лексико-семантические преобразования, но и синтаксические. Для троек русско-немецко-английских патентных параллельных текстов, в свою очередь, наиболее характерными оказались *предложно-номинативные* и *адъективно-номинативные* трансформации, которые также являются самыми частотными для русско-немецких параллельных текстов. При этом для всех типов текстов в русскоязычных формулах изобретения характерно превалирова-

ние назывных конструкций с распространениями, в то время как в немецкоязычных текстах доминируют глагольные структуры (на всех уровнях анализа предложения).

Благодарности

Автор выражает благодарность Елене Борисовне Козеренко за конструктивное обсуждение данной работы.

Литература

1. Козеренко Е. Б. Лингвистическое моделирование для систем машинного перевода и обработки знаний // Информатика и ее применения, №1, том 1. — М.: Торус, 2007. — С.54–65.
2. Козеренко Е. Б. Глагольно-именные трансформации при англо-русском машинном переводе // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2007» / Под ред. Л. Л. Иомдина, Н. И. Лауфер, А.С. Нариньяни, В. П. Селегея. — М.: Изд-во РГГУ, 2007. — С. 286–294.
3. Жолковский А. К., Мельчук И. А. О семантическом синтезе «Проблемы кибернетики», вып. 19. М, 1967.
4. Падучева Е. В. О семантике синтаксиса. Материалы к трансформационной грамматике русского языка. Изд. 2-е. М: КомКнига, 2007, 296 с.
5. Jacobs, Roderick A. and Peter S. Rosenbaum. English Transformational Grammar. Blaisdell, 1968.
6. Балли, Ш. Общая лингвистика и вопросы французского языка. Изд. 2-е, М.: УРСС, 2001.
7. Кузнецов И. П., Мацкевич А. Г. Семантико-ориентированные системы на основе баз знаний. — М.: из-во МТУСИ, 2007. — 173 с., 59 ил.
8. Козеренко Е. Б. Проблема эквивалентности языковых структур при переводе и семантическом выравнивании параллельных текстов // Компьютерная лингвистика и интеллектуальные технологии: Труды международной конференции «Диалог 2006» / Под ред. Л. Л. Иомдина, Н. И. Лауфер, А. С. Нариньяни, В. П. Селегея. — М.: Изд-во РГГУ, 2006. — С. 252–258.
9. Nivre J., Boguslavski I., Iomdin L. Parsing the SynTagRus Treebank of Russian \ Proceedings of the International Conference COLING'2008, Manchester, UK, 2008.
10. Macken L., Lefever E., Hoste V. Linguistically-based sub-sentential alignment for terminology extraction from a bilingual automotive corpus \ Proceedings of the International Conference COLING'2008, Manchester, UK, 2008.