

УДК 811.112.2'322.4

Детскина Раиса Владимировна

кандидат филологических наук, доцент,
доцент кафедры информатики
и прикладной лингвистики
Минский государственный
лингвистический университет
г. Минск, Беларусь

Raisa Detskina

PhD in Philology, Associate Professor
Associate Professor of the Chair
of Informatics and Applied Linguistics
Minsk State Linguistic University
Minsk, Belarus
raisadetskina@gmail.com

СТРУКТУРНО-СИНТАКСИЧЕСКИЕ ПРЕОБРАЗОВАНИЯ
ПРИ МАШИННОМ НЕМЕЦКО-РУССКОМ ПЕРЕВОДЕ
НАУЧНО-ТЕХНИЧЕСКИХ ТЕКСТОВ

STRUCTURAL AND SYNTACTIC TRANSFORMATIONS
IN THE AUTOMATIC GERMAN-RUSSIAN TRANSLATION
OF SCIENTIFIC TECHNICAL TEXTS

В данной статье рассматриваются структурно-синтаксические преобразования при автоматическом немецко-русском переводе научно-технических текстов. В процессе анализа были выделены структурные типы моделей немецких заголовков рефератов к патентам и модели их перевода на русский язык. Даны примеры, иллюстрирующие полученные в ходе анализа результаты.

The article considers structural and syntactic transformations of scientific technical texts in the process of automatic German-Russian translation. Structural model types of German headings to abstracts of patents have been presented; ways of translating them into Russian have been defined in the research. The obtained results are proved by examples.

К л ю ч е в ы е с л о в а: структурно-синтаксические преобразования; машинный перевод; научно-технический текст; немецко-русский перевод.

Key words: structural and syntactic transformations; automatic translation; scientific technical text; German-Russian translation.

Материалом для анализа в данной работе послужили 50 немецких заголовков рефератов к патентам по тематикам «Моделирование голоса, использование техники моделирования голоса при распознавании речи», «Синтезирование речи, способы», «Распознавание личности говорящего по речи».

В исследуемых заголовках имя существительное образует ряд синтаксических моделей в результате пре- и постпозитивного распространения именного ядра заголовка. Мы именуем их базовыми моделями. *Базовые синтаксические модели* – это n-компонентные словосочетания, включающие только полные слова. Формальными разделителями базовых моделей служат, главным образом, предлоги, союзы, знаки пунктуации, а также отдельные словоформы.

При выделении структурных типов моделей целесообразно исходить из степени спаянности и характера синтаксических отношений между составляющими компонентами модели, число которых равно двум. Следовательно, характерной чертой соединения слов в заголовке является то, что они создают бинарную структуру. Кроме моделей, включающих одиночное ядерное существительное или несколько существительных, соединенных союзами *und* или *oder*, любая базовая синтаксическая модель состоит из главного (определяемого) и одного или нескольких зависимых элементов (определяющего). Например, *Virtuelles Testbett* ‘способ опроса сети’ (DE 10035523 A1)¹.

Даже если базовая синтаксическая модель содержит три и более члена, например, *Automatische Signalpegelanpassung für Spracherkennungssysteme* ‘способ автоматической настройки уровня сигнала для систем распознавания речи’ (DE 10209323 C1), в силу линейного характера заголовков все члены таких конструкций образуют группирующиеся компоненты, что позволяет выделить несколько структурных типов, хотя общая структура остается бинарной. При этом вполне закономерно, что словоформы, зависящие от ядерной словоформы (существительного), могут, в свою очередь, иметь при себе подчиненные словоформы второго ранга, а последние, соответственно – словоформы третьего ранга и т. д. Например, *System und Verfahren zur parallelen Darbietung von mehreren Audioinformationsquellen* ‘система и способ параллельной передачи аудиоинформации’ (DE 10123823 A1). В приведенном примере словоформы *System* ‘система’, *Verfahren* ‘способ’ имеют подчиненную словоформу *Darbietung* ‘передача’, которая в свою очередь имеет свою подчиненную словоформу *Audioinformationsquelle* ‘аудиоинформация’ (подчинение второго ранга).

Элементы языка обладают определенными синтагматическими сочетательными возможностями. И эти особенности различны у элементов, входящих в рассматриваемые нами базовые синтаксические модели. Здесь имеются в виду, с одной стороны, существительные, а с другой – прилагательные, причастия, местоимения и числительные. Они и определяют образования или невозможность образования определенных группировок элементов.

В анализируемых заголовках рефератов к патентам простейшей структурой является **S** (существительное). Например, *Sprachfenster* ‘способ вывода речевой информации’ (DE 10143292 A1), *Sprachdialogsystem* ‘система речевого диалога’ (DE10043531 A1).

К тому же в исследованных заголовках имеется ряд случаев, где ядерное существительное имеет левые и правые распространители и их комбинации. В зависимости от синтаксического типа распространителей и их позиции по отношению к ядру, выделен ряд структурных вариантов ядерной модели **S** (примеры таких моделей представлены в табл. 1).

¹ Здесь и далее по тексту в скобках указаны выходные данные рефератов к патентам: код страны, номер документа, вид документа.

При анализе были выделены и двухкомпонентные базовые синтаксические модели. Им, в первую очередь, свойственна бинарность отношений между их компонентами. В данной модели левый одиночный распространитель может быть выражен прилагательным, причастием или местоимением (модель **A + S**). Например, *Virtuelles Testbett* ‘способ опроса сети’ (DE 10035523 A1). Хотя данная модель и не является частотной в проанализированных заголовках (13,7 %), все же ее количественный показатель выше по сравнению с другими двухкомпонентными структурами. Причем в большинстве случаев эта модель не является одиночной при номинации патента. В одном и том же заголовке могут встречаться и другие модели. Например, *Automatische Erkennung von Unternehmensnamen in sprachlichen Äußerungen* ‘способ автоматического распознавания названий компаний в фразах речи’ (DE 10038517 A1).

Правый одиночный распространитель для двухкомпонентных базовых моделей может быть выражен в том числе существительным (модель **S₁ + S₂**). Например, *Verfahren zum Betrieb eines Sprach-Dialogsystems* ‘способ работы речевой диалоговой системы’ (DE 10156945 A1). Их процентное представление в массиве заголовков также невелико (5,8 %).

Трехкомпонентные базовые синтаксические модели образуются, главным образом, в результате распространения определения, выраженного прилагательным, причастием или местоимением. Они выступают в качестве левого распространителя для ядерного существительного. Однако в представленных заголовках распространение определения, выраженного местоимением, встретилось лишь один раз, что является нетипичным для языка немецкой научной литературы в целом. А именно: *Zuordnen von Phonemen zu den sie erzeugenden Graphemen* ‘способ соотнесения фонем с формируемыми графемами’ (DE 10042943 A1). В данном примере выраженное причастием I прилагательное и местоимение, к нему относящееся, образуют причастный оборот. Тем самым между артиклем и существительным формируется рамка.

Четырех-, пяти-, шестикомпонентные модели образуются путем распространения выделенных трехкомпонентных моделей. В проанализированных заголовках представлена лишь одна четырехкомпонентная синтаксическая модель, которая составляет незначительную часть рассмотренного массива (1,5 %). Например, *Intelligente Korrektur diktiert Sprache* ‘способ и система для автоматической корректировки диктуемой речи’ (DE 10040214 A1).

Как показал анализ, заголовки рефератов к патентам чаще всего представляют собой комбинации нескольких моделей. Например, **(S zur S von S in S)** *System zur Umsetzung von Textdaten in eine Sprachausgabe* ‘система для преобразования текста в речь’ (DE 10160673 A1).

Следует подчеркнуть, что в большинстве из исследуемых заголовков рефератов к патентам представлена структурная модель **S**. Заголовки носят номинативный характер, что является характерным для научного стиля. Усложнение немецких заголовочных конструкций происходит за счет однород-

ных членов, вводимых союзами, а также вследствие того, что в большинстве случаев заголовки немецких рефератов к патентам представляют собой комбинацию нескольких базовых синтаксических моделей.

В табл. 1 представлены выделенные базовые модели, их примеры и частота употребления в исследуемом массиве. При этом статистические данные учитывают встречаемость моделей в случае их комбинаций в одном и том же заголовке.

Т а б л и ц а 1

Базовые синтаксические модели
проанализированных немецких заголовков рефератов к патентам

№	Структурная модель	Примеры	Частота употребления, %
1	S	<u>Stimmenkonverter</u> 'устройство для преобразования речи'	73,9
2	S ₁ + S ₂	Verfahren zum <u>Betrieb eines Sprach-Dialogsystems</u> 'способ для работы речевой диалоговой системы'	5,8
3	A + S	Vorrichtung und Verfahren zum <u>graphischen Darstellen</u> und zum Übertragen von Informationen 'способ и устройство для графического представления и передачи информации'	13,7
4	Pron + S	System und Verfahren zur parallelen Darbietung von <u>mehreren Audioinformationsquellen</u> 'система и способ параллельной передачи объемной аудиоинформации'	0,72
5	S ₁ + A + S ₂	<u>Spracheingabe gesteuertes Steuergeräts</u> 'устройство управляемого речевого прибора'	2,2
6	S ₁ + (und-) + S ₂	<u>Lautverarbeitungsvorrichtung und -verfahren</u> 'способ и устройство для обработки звука'	0,72
7	Pron + A + S	Zuordnen von Phonemen zu den <u>sie erzeugenden Graphemen</u> 'способ соотнесения фонем с формируемыми их графемами'	0,72
8	S ₁ + S ₂ (und) + S ₃	Verfahren zur <u>Steuerung einer Spracheingabe und -ausgabe</u> 'способ для управления вводом речи и выводом'	0,72
9	A ₁ + S ₁ + A ₂ + S ₂	<u>Intelligente Korrektur diktierter Sprache</u> Erkennung einer in <u>buchstabierter Form vorliegenden Sprachäußerungseingabe</u> 'распознавание в буквенной форме представленных фраз речи'	1,52

Следующим этапом данной работы является установление типологии базовых синтаксических моделей немецких заголовков и их русских соответствий как основы алгоритма машинного перевода.

Для разработки алгоритма немецко-русского перевода заголовков были выявлены структурные соответствия немецких заголовков, образующих базовые модели, в русском языке. Сравнительный анализ структур заголовков показал, что для большинства базовых моделей немецких заголовков существует однозначное соответствие их структур в русском языке (табл. 2).

Т а б л и ц а 2

Структурные соответствия
немецких синтаксических моделей в русском языке

№	Немецкие базовые синтаксические модели и их примеры	Структурные типы русских переводов и их примеры	Употребительность выделенных моделей в исследуемом массиве заголовков, %
1	S <i>Stimmenkonverter</i>	s <u>Устройство для преобразования речи</u>	73,9
2	S₁ + S₂ <i>Verfahren zum Betrieb eines Sprach-Dialogsystems</i>	s ₁ + s ₂ (р.п.) <u>Способ для работы речевой диалоговой систем</u>	5,8
3	A + S <i>Automatische Signalpegelanpassung für Spracherkennungssysteme</i>	1. a + s <u>Автоматический способ настройки сигнала для систем распознавания речи</u>	13,7
4	Pron + S <i>System und Verfahren zur parallelen Darbietung von mehreren Audioinformationsquellen</i>	a + s <u>Система и способ параллельной передачи объемной аудиоинформации</u>	0,72
5	S₁ + (und-) + S₂ <i>Lautverarbeitungsvorrichtung und -verfahren</i>	s ₂ + (и) + s ₁ <u>Способ и устройство для обработки звука</u>	0,72
6	S₁ + A + S₂ <i>Spracheingabe gesteuertes Steuergeräts</i> <i>Verfahren und Vorrichtung zum Erkennen einer phonetischen Lautfolge oder Zeichenfolge</i>	1. s ₁ + a (р.п.) + s ₂ (р.п.) <u>Устройство управляемого речевого прибора</u> 2. s ₁ + (a) + s ₂ => s ₁ + + s ₂ <u>Способ и устройство для распознавания последовательности фонем или последовательности знаков</u>	1,45 0,75

Окончание таблицы 2

7	Pron + A + S <i>Zuordnen von Phonemen zu den <u>sie</u> erzeugenden Graphemen</i>	a + pron (р.п.) + s Способ соотнесения фонем с <u>формируемыми их графемами</u>	0,72
8	S₁ + S₂ + (und) + S₃ <i>Verfahren zur <u>Steuerung einer</u> Spracheingabe und -ausgabe</i>	s ₁ + s ₂ (тв.п.) + (и) + s ₃ (тв.п.) Способ для <u>управления вводом</u> речи и <u>выводом</u>	0,72
9	A₁ + S₁ + A₂ + S₂ <i>Erkennung einer in <u>buchstabierter Form vorliegenden</u> Sprachäußerungseingabe</i>	a ₁ + s ₁ + a ₂ (р.п.) + s ₂ (р.п.) Распознавание в <u>буквенной</u> форме <u>представленных фраз</u> речи	1,52

На основе установленной типологии базовых синтаксических моделей немецких заголовков и их русских соответствий создан принципиальный алгоритм немецко-русского машинного перевода заголовков рефератов к патентам.