

перспективности они не способны полностью заменить переводчика и сами по себе не гарантируют качественного перевода. Как отмечает В.Н. Шевчук, «качественный перевод возможен лишь при разумном сочетании новых информационных технологий с наработанными переводческими приемами и навыками, хорошим знанием иностранного языка и теории перевода, данных сопоставительной лингвистики в конкретной языковой комбинации и большим практическим опытом» [8, с. 335]. Поэтому, в конечном счете, все зависит от личности переводчика, его профессионального опыта, навыков и находчивости. Информационные технологии и электронные ресурсы являются лишь полезным, а иногда и необходимым дополнением к его творчеству.

### Библиографические ссылки

1. *Austermühl F.* Electronic Tools for Translators: Translation Practice Explained / F. Austermühl. London; N.Y.: Routledge, 2014. P. 202.
2. *Zampieri M.* Quantifying the Influence of MT Output in the Translators' Performance: a Case Study in Technical Translation / M. Zampieri, M. Vela // Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation. Association for Computational Linguistics. Sweden, 2014. P. 93–98.
3. *Груздев Д.Ю.* Электронный корпус текстов как эффективный инструмент переводчика: автореф. дис. ... канд. филол. наук: 10.02.19 / Д.Ю. Груздев. М. : Изд-во Военного ун-та, 2013. 20 с.
4. *McEnery T.* Corpus Linguistics / T. McEnery, A. Wilson. Edinburgh: Edinburgh Univ. Press, 2001. 209 p.
5. *Андреева Е.Г.* Анализ переводческих соответствий на материале параллельного корпуса текстов / Е.Г. Андреева // Компьютерная лингвистика и интеллектуальные технологии: труды междунар. конф. «Диалог 2006» (2006 ; Бекасово) / под ред. Н.И. Лауфер, А.С. Нариньяни, В.П. Селегея. М. : Изд-во РГГУ, 2006. С. 26–30.
6. *Добровольский Д.О.* Корпус параллельных текстов и литературный перевод / Д.О. Добровольский // НТИ. Сер. 2. 2003. № 10. С. 13–18.
7. *Владимов Н.В.* Корпусный подход к решению переводческих проблем: автореф. дис. ... канд. филол. наук: 10.02.19 / Н.В. Владимов. М.: Изд-во Военного ун-та, 2005. 20 с.
8. *Шевчук В.Н.* Информационные технологии в переводе. Электронные ресурсы переводчика / В.Н. Шевчук. М.: Зебра-Е, 2013. 384 с.

*Н.А. Метлицкая (Минск, Беларусь)*

### ПРИНЦИП РАБОТЫ АЛГОРИТМА АВТОМАТИЧЕСКОГО ПОРОЖДЕНИЯ АНГЛОЯЗЫЧНЫХ РЕКЛАМНЫХ ТЕКСТОВ

В статье представлено описание принципа работы алгоритма системы автоматического порождения англоязычных рекламных текстов по косметике и

парфюмерии. Разработанный алгоритм определяет порядок следования операций, выполняемых при планировании содержания текста, а также при выборе лексических и грамматических средств, представленных в лингвистической базе знаний системы автоматического порождения, и их оформлении в связный текст.

**Ключевые слова:** *автоматическое порождение текста, лингвистическая база знаний, формальный язык, электронный словарь лексических единиц, алгоритм автоматического порождения текста.*

Среди современных систем автоматического порождения текста можно наблюдать два принципиально различных типа систем: системы, основанные на использовании шаблонов (шаблонные системы), и системы, основанные на использовании лингвистической базы знаний (лингвистически мотивированные системы).

Первые предполагают использование при синтезе текста готовых текстовых фрагментов, заложенных в базе данных – слов, предложений, абзацев. По сути, как такового «порождения» или «синтеза» нового текста в данном случае не происходит, т.к. система использует готовые текстовые фрагменты, которые передаются на экран компьютера при определенном запросе пользователя. Системы второго типа работают с содержанием будущего текста, представленном в виде данных нетекстовой природы – базы знаний, семантические и формальные языки [1]. Принцип работы таких систем подразумевает моделирование процесса порождения текста человеком, то есть создание текста происходит от некоторого концептуального уровня (зарождение идеи, замысла) к текстовому (их языковому оформлению). Практическая реализация лингвистически мотивированных систем предполагает наделение их широким спектром лингвистических знаний как о структуре, так и о содержании порождаемого текста, а также сложными лингвистическими знаниями, которые позволяют выразить это содержание языковыми средствами. Ведь создаваемый текст должен быть корректным на всех языковых уровнях – грамматическом, лексическом, семантическом. Иными словами, в процессе работы по созданию текста лингвистически мотивированные компьютерные системы опираются на обширную, сложную и многокомпонентную лингвистическую базу знаний, а согласованная и координированная работа всех ее компонентов обеспечивается четкими предписаниями алгоритма.

С целью реализации лингвистически мотивированного подхода при создании системы автоматического порождения англоязычных рекламных текстов нами была разработана лингвистическая база знаний, в которую вошли следующие компоненты:

- 1) списки главных и второстепенных опорных слов (ГОС и ВОС) исследованных рекламных текстов;
- 2) словарь лексических единиц с указанием семантических и

морфологических сведений о каждой лексической единице;

3) семантико-синтаксические формулы текстов на формальном языке СЕМСИНТ.

Далее был разработан алгоритм автоматической системы порождения рекламных текстов, который определяет порядок следования операций, выполняемых при планировании содержания текста, а также при выборе лексических и грамматических средств, представленных в лингвистической базе знаний, и их оформлении в связный текст [2].

В самом общем виде порождение текста в соответствии с данным алгоритмом предполагает последовательное решение следующих задач (рис. 1):

1. Определение темы рекламного сообщения

2. Определение семантико-синтаксического содержания будущего текста (выбор семантико-синтаксической формулы текста)

3. Лексическое оформление семантико-синтаксической формулы текста

4. Морфологическое оформление слов текста



Рис. 1. Принципиальная схема автоматического порождения рекламного текста

Представленная на рисунке 1 схема наглядно демонстрирует, что на всех основных этапах по созданию текста автоматическая система постоянно взаимодействует с лингвистической базой знаний.

Рассмотрим порядок действий алгоритма автоматического порождения рекламного текста на конкретном примере.

Работа алгоритма начинается с выведения на экран возможных тем рекламного сообщения, которые может создать система порождения: губная помада, тушь для ресниц, шампунь. Пользователю предлагается выбрать одну из трех тем. Предположим, пользователь запросил создать рекламный текст на тему «Шампунь».

После того как выбрана тема рекламного текста, начинается последовательная работа системы по планированию и оформлению его содержания. Модели порождаемых рекламных текстов представлены в виде

семантико-синтаксических формул текстов на формальном языке СЕМСИНТ [3]. Система обращается к области базы знаний, содержащей семантико-синтаксические формулы текстов на заданную тему («Шампунь»), и датчик случайных чисел выбирает одну (любую) из десяти соответствующих данной теме формул. Тем самым обеспечивается разнообразие и вариативность семантико-синтаксической структуры порождаемых системой рекламных текстов, а также обеспечивается учет вероятностных факторов, которые играют важную роль при порождении текста человеком.

Пусть на данном этапе системой выбрана семантико-синтаксическая формула текста S01:

АН1<A52/вос1\*N75/1/roc1.02\*N89/1>.

R<V32/2> + АО<A52/вос1\*N50/2>.

R2<V20/2> + АО<N31/1> + AAD<J03\*T01\*A58\*N75/1/roc1.01>.

R2<V06/1> + АО<P10\*N98/1> + AAD<J03\*T02\*N03/1> + АН1<K04\*A33\*N11/1/вос2> + R<V47/1\*V16/1>!

R<V07/4> + АКМ<J03\*A52/вос1\*N50/2>,

АН1<A52/вос1\*N75/roc1.02\*N89/1> + R<V22/2> + АО<A36\*

N11/1/вос2\*C01\*N13/1> + AAD<J03\*T01\*A58\*N75/1/roc1.02>.

R2<V07/4> + АКМ<J03\*A21\*N75/1/roc1.01\*N83/1> + PPS<J05\*T01\*A69\*N75/1/roc1.01\*N35/1>,

R<V37/2> + АО<P10\*N11/1/вос2> + AD1<V38/3\*A04,

\*V40/3\*A68\*C01\*100%\*A33> + АТМ<P15\*N98/1\*S15>.

Затем, в соответствии с предписанием алгоритма, осуществляется лексическое заполнение выбранной семантико-синтаксической формулы текста. На данном этапе система взаимодействует одновременно с двумя областями базы знаний – списками главных и второстепенных опорных слов и электронным словарем. Каждому тексту соответствует свой набор главных и второстепенных опорных слов, которые отмечены в семантико-синтаксической формуле данного текста и заполняют, таким образом, конкретные позиции в формуле (тексте). Если слово не относится к ГОС или ВОС, система производит обращение к электронному словарю, из которого с учетом семантических кодов слов, указанных в формуле, выбираются соответствующие лексические единицы. Так, при прочтении первого закодированного слова – A52/вос1 – система определяет, что выделенное слово относится к ВОС1. Следовательно, система обращается к области базы знаний, в которой содержатся списки ВОС текста S01, и берет слово, соответствующее слову ВОС1. В данном случае ему соответствует слово *citrus* ‘цитрусовый’. Далее последовательная работа системы продолжается в блоке 4, когда выбранное слово должно быть оформлено в соответствии с морфологической информацией, указанной о нем в формуле. Код A52

указывает на то, что это имя прилагательное, относящееся к семантическому подклассу «прилагательные, характеризующие объект по отношению к запаху». Грамматическая (морфологическая) информация об имени прилагательном в формуле и словарной статье не указана. Следовательно, данное слово *citrus* подставляется в формулу без каких-либо преобразований. При прочтении очередного закодированного слова N75/1/гос1.02 система определяет, что оно относится к ГОС1.02, которое также содержится в соответствующей области базы знаний. Однако данная запись – N75/1/гос1.02 – содержит грамматическую информацию. А именно, обозначение «/1» после семантического кода N75 свидетельствует о том, что данное слово в формуле является именем существительным в единственном числе. Следовательно, на этапе работы по морфологическому оформлению лексических единиц текста (блок 4) система выберет из электронного словаря имя существительное с семантическим кодом N75 в форме единственного числа. То есть, в данном месте в формуле будет стоять одно из следующих слов: *scent* ‘запах’, *fragrance* ‘аромат’ или *odour* ‘аромат’ в форме единственного числа.

Аналогичная ситуация наблюдается с очередным закодированным словом N89/1, которое также является именем существительным в единственном числе. Немного иная картина будет наблюдаться, когда в процессе дальнейшей работы система прочтет, например, закодированное слово N50/2 во втором предложении. Запись «/2» после семантического кода N50 означает, что данное существительное должно быть употреблено во множественном числе, поэтому при обращении к электронному словарю система возьмет из него форму имени существительного с семантическим кодом N50 во множественном числе. Следовательно, в порожденном предложении на этом месте будет стоять либо слово *properties*, либо слово *essences* ‘эссенции’, которые имеют данный семантический код.

Таким образом, в процессе лексического заполнения семантико-синтаксической формулы текста система последовательно «читает» каждое закодированное слово и прежде всего определяет, является ли очередное прочитанное слово главным или второстепенным опорным словом? Если закодированное слово относится к ГОС или ВОС, система порождения обращается к соответствующей области базы знаний, чтобы извлечь оттуда конкретное слово. Если прочитанное слово не относится к ГОС или ВОС, система производит обращение к электронному словарю, из которого с учетом семантических кодов слов, указанных в формуле, выбираются соответствующие лексические единицы.

Далее, на этапе грамматического оформления предложений генерируемого текста осуществляется морфологическое оформление лексических единиц, которые были выбраны системой для заполнения

данной семантико-синтаксической формулы. Необходимые морфологические преобразования совершаются в процессе автоматического порождения текста с именами существительными и глаголами в соответствии с кодами грамматических форм этих лексических единиц, указанных в формуле, а также в соответствии с грамматическими сведениями о данных лексических единицах, которые представлены в словаре базы знаний системы. Если закодированные слова имеют коды S, P, J, C, M, K или T, никаких морфологических преобразований не требуется, так как данные коды соответствуют наречиям, местоимениям, предложениям, союзам, числительным, частицам и артиклям, которые являются неизменяемыми частями речи. Прочитав очередное закодированное слово, имеющее один из перечисленных кодов, система каждый раз обращается к соответствующей области базы знаний и выбирает слово с необходимым семантическим кодом. Например, P10, T01, J03 и т.д. Каждый из таких кодов обозначает конкретное местоимение, артикль, предлог и т.д. Дополнительная морфологическая информация не указывается также для имен прилагательных. Объясняется это тем, что единственной грамматической категорией имен прилагательных в английском языке является степень сравнения. Поскольку в исследованных рекламных текстах, которые легли в основу базы знаний описываемой системы, все имена прилагательные употреблены в положительной степени сравнения, в указании дополнительных сведений о формах других степеней необходимости нет.

Таким образом, разработанный алгоритм обеспечивает последовательное – по предложениям – прочтение и заполнение системой выбранной семантико-синтаксической формулы. Когда очередное предложение окончено, начинается работа по созданию нового предложения текста. После того, как вся семантико-синтаксическая формула оказывается заполненной конкретными словами в необходимой морфологической форме, готовое рекламное сообщение выводится на экран компьютера.

### **Библиографический список**

1. *Соколова Е.Г.* Автоматическая генерация текстов на ЕЯ (портрет направления). URL: <http://www.dialog-21.ru/archive/2004/sokolova.htm> (дата обращения: 08.04.2015).
2. *Метлицкая Н.А.* Лингвистическая база данных системы автоматического порождения англоязычного рекламного текста // Системный анализ и прикладная информатика. Минск : БНТУ, 2017. № 2. С.62–67.
3. *Зубов А.В.* Семантико-синтаксический язык для записи текстов в памяти ЭВМ // Функционирование и развитие языковых систем : сб. науч. тр. Минск : Вышэйшая школа, 1990. С. 110–117.