

Автоматическая генерация текстов на ЕЯ (портрет направления)

Е. Г. Соколова

Российский НИИ искусственного интеллекта
125190, Москва, а/я 85
e-mail: sokolova@aha.ru

М. В. Болдасов

Российский НИИ искусственного интеллекта
125190, Москва, а/я 85
e-mail: boldasov@mn.ru

Автоматическая Генерация на Естественном Языке (ГЕЯ) очень слабо представлена в отечественной науке последних лет. В статье коротко описывается состояние дел в этой области по материалам в основном зарубежной печати и Интернет с упором на опыт создания прикладных систем генерации монологических текстов. Обсуждаются виды и свойства входных данных для таких систем, типы и свойства выходных текстов, а также основные процессы генерации, определенные соотношением входных данных и типами выходных текстов.

1. Введение

В настоящее время письменный ЕЯ текст широко распространен как средство коммуникации пользователя с компьютерными системами. Большая часть программных приложений, которые выдают тексты на ЕЯ, работает с ЕЯ информацией как со строкой символов. Они манипулируют готовыми предложениями и словосочетаниями как строительными блоками будущего текста. Это - шаблонные технологии. Шаблонные технологии относительно просты и надежны и находят широкое промышленное применение. Главная особенность этих технологий состоит в том, что содержание будущего текста представлено в них в виде фрагментов текста.

Другой вид систем работает с содержанием будущего текста, представленном в виде данных нетекстовой природы (БД, баз знаний, семантических и формальных языков). В этом случае для создания текста системе необходимы знания структуры содержания и знания об устройстве генерируемого текста, а также сложные лингвистические знания, которые позволяют выразить это содержание языковыми средствами.

Это - Лингвистически Мотивированные (ЛМ) технологии. В настоящее время они еще находятся в стадии исследования. ЛМ технологии связаны с изучением жанровых и коммуникативных характеристик текстов, риторических приемов организации содержания

текста, языковых средств выражения связности текста (лингвистика текста), формализацией грамматики и лексических описаний.

Мотивации выборов языковых средств выражения знаний в ЛМ технологиях подобны мотивациям движений летящей птицы. Они не могут быть буквально повторены автоматической системой. Поэтому, хотя ЛМ технологии опираются на определенные соответствия неязыковых знаний и языковых средств, они также активно используют и шаблонные технологии для решения конкретных проблем.

2. Шаблонные системы

Шаблонная система использует готовые реплики или комбинирует готовые фрагменты текста таким образом, что они занимают заданные позиции в дискурсе или стереотипном тексте. Самые простые шаблонные системы просто вставляют фрагменты текста в шаблоны без их дополнительной обработки, например, реплика системы: «Не могу найти my1.txt, my2.txt файл(ы)!»..

Более сложные шаблонные системы дополнительно проводят ограниченную лингвистическую и риторическую обработку результата - позволяют задавать отдельные грамматические параметры текста или комбинировать шаблонные высказывания в связный текст, используя определенные лексические и грамматические знания о ЕЯ. К последнему типу относятся, в частности, системы, помогающие менеджерам писать деловые характеристики на сотрудников. Например, система Employee Appraiser (производитель Austin-Haynes), Performance Now (производитель KnowledgePoint). Такие системы предоставляют пользователю набор оценочных тем, таких как *Общение*, которые, в свою очередь, разбиваются на более специфичные подтемы, например, *Устный обмен идеями*. Выбирая свойство, подходящее для характеристики данного работника, менеджер автоматически выбирает и конкретный готовый абзац или предложение, которое система сама вставляет в результирующий текст. При этом система может использовать сведения о поле сотрудника (для грамматического согласования прилагательных и глаголов с именами), указать, в каком лице должен быть составлен текст деловой характеристики (лицо глаголов), осуществлять простейшее планирование текста и предложений. Например, объединить всю информацию об отдельной высокоуровневой теме (такой, как *Общение*) в отдельный абзац, используя сочинительные конструкции и местоимения (агрегация предложений), а также вставку слов-коннекторов в начале предложения. В следующем примере на рис. 1 подчеркнуты слова, вставленные системой для создания связности текста:

Bert does not display the verbal communication skills required, and his written communications fall short of the quality needed. Additionally, he does not exhibit the listening and comprehension skills necessary for satisfactory performance of his job.

Поскольку содержание в шаблонных системах представлено непосредственно в виде фрагментов текста, порожденные ими тексты выглядят абсолютно естественными, но они работают с очень жесткими типами текстов.

3. Системы автоматической генерации текстов на ЕЯ (ГЕЯ)

Генерация на естественном языке (ГЕЯ) как научное направление занимается созданием компьютерных систем, производящих тексты на естественном языке (ЕЯ) из некоторого нелингвистического (нетекстового) представления информации. Первые системы в доступном в Интернете каталоге систем ГЕЯ (The John Bateman and Michael Zock's list of Natural Language Generation Systems), содержащем краткую информацию о всех известных авторам каталога системах ГЕЯ относятся к 63-65 гг. Всего каталог содержит данные по

более чем 340 системам автоматической генерации текстов с начала исследований по настоящее время.

ГЕЯ активно развивается за рубежом начиная с 90ых годов, прежде всего в США, Канаде и странах Западной Европы - в Великобритании, Германии, Голландии и др., а также в Австралии. В настоящее время ГЕЯ представляет собой динамичное направление АОТ. В Ассоциации по компьютерной лингвистике ACL создана группа по интересам, объединяющая ученых и коллективы, работающие в области ГЕЯ –Special Interest Group on GENERation (SIGGEN). Раз в два года под эгидой SIGGEN проводятся научные конференции INLG, посвященные ГЕЯ.

ЛМ системы предназначены для создания текстов, имеющих относительно свободное содержание, которое не может быть заранее задано в виде готовых фрагментов текста. Источником содержания являются данные, представленные в виде БД, БЗ, или в виде выражений на формализованных языках, например, SQL. Тип входных данных не всегда предсказывает тип выходного текста, поэтому тип текста определяется из-вне. Часто ЛМ система создает тексты одного типа, но в разных ПО, или в одной ПО, но на разных языках. Несмотря на сложность ЛМ систем, они имеют важные преимущества по сравнению с использованием ручного труда и шаблонными системами. Эти преимущества обсуждались в ряде работ, например, в (Reiter, 1995). К ним относятся: *сопровождаемость*, лучшее *качество создаваемых текстов*, *многоязыковой выход* и *гарантированное соответствие стандартам*.

3.1. Организация системы ГЕЯ: ресурсы и обрабатывающий компонент

Программный код, реализующий генератор обычно разделяется на два компонента: *ресурсы* и *обрабатывающий компонент*. Ресурсы описывают знания, необходимые для генерации, а обрабатывающий компонент применяет эти знания к входным данным. Для многоязыковых генераторов строится единый обрабатывающий компонент, реализующий некоторую лингвистическую идеологию, а для выходных языков создаются ресурсы - языковые средства для конкретного ЕЯ – грамматики, словари. Примером обрабатывающего компонента может служить среда для создания генераторов KPML (KPML). Она использовалась в нескольких проектах, в частности, в проекте генерации инструкций к графическим редакторам на болгарском, русском и чешском языках AGILE (Kruijff et al., 2000), (Соколова, Шаров, 1998). В AGILE создается единое для трех языков представление содержания текста – A-box (assertion box), оно разбивается на семантические представления отдельных предложений, которые имеют локальные отличия по выходным языкам.

Языковые ресурсы обычно модуляризируются - разделяются по уровням лингвистического описания и по специфичности описания для конкретного языка. Традиционно выделяются следующие уровни лингвистического описания: *лексика*, *морфология*, *синтаксис*, *семантика*, *риторика* и *прагматика*. Первые три уровня описывают явления, специфичные для конкретного языка. Вторые три уровня задают межязыковые знания, структуру текста и его контекст. Разделением ресурсов генерации на модули согласно уровням лингвистического описания достигается последовательность в ограничении области поиска альтернатив в процессе генерации.

Ресурсы должны быть максимально независимы от особенностей конкретного приложения, а все особенности задачи должны быть отражены в принимаемых генератором на входе данных. В этом случае задачу построения генератора необходимо решить только один раз для каждого описываемого ЕЯ. К настоящему моменту считается, что генератор, реализующий такую схему, на входе должен принимать *систему знаний*, из которой будет конструироваться текст, *коммуникативную цель* порождаемого текста, *модель адресата* текста, и *контекст повествования* (Reiter and Dale 2000).

3.2. Условная схема процесса генерации

Общая и полная схема генерации без детализации происходящих процессов состоит из трех основных блоков:

Планирование содержания текста Микропланирование Реализация на ЕЯ
Планирование содержания текста – решение, какая именно информация из входных данных попадет в текст, и как она будет организована. Этот этап работает исключительно с предметным знанием и общими способами организации содержания в тексте. Результатом является план текста в терминах последовательности событий, метафункций, например, *запрос информации, предоставление информации*, или риторических отношений.

Микропланирование – это интерфейсный блок, который позволяет от предметных знаний перейти к языковым. В нем решается, каким образом выбранная информация будет реализована языковыми средствами в виде предложений на ЕЯ. Результатом этого процесса являются представления предложений в виде структур семантических и/или синтаксических отношений.

Реализация на ЕЯ – производство грамматически правильных предложений текста. Основан на лингвистических знаниях. Этот блок часто выделяется как универсальный и включает в себя либо только морфологический синтез словоформ, либо переход от семантических представлений к поверхностно синтаксическим и синтез словоформ, например, ресурсы конкретных языков в AGILE.

В (Bateman and Zock, 2003) рассматривается более подробная схема процесса генерации. Схема генерации описывает некий идеализированный процесс. Действующая модель генерации не является и не может быть полной и последовательной реализацией схемы генерации. Разработчики постоянно сталкиваются с проблемами, которые получили название *generation gaps* (“пропасти”), которые на практике преодолеваются директивным сопоставлением средств выражения (Jokinen et al. 1996). Таким образом, вместо плавного уровневого процесса генерации в действующей системе мы имеем рваную модель, достигающую своей цели отчасти эмпирическими шаблонными методами, применяемыми локально. Тем не менее, описание обобщенной картины процесса генерации от концептов исходных данных до ЕЯ текста имеет важное методологическое значение.

3.3. Входные данные и системы ГЕЯ

Определение вида входных данных является кардинальным вопросом для ЛМ систем. Теоретическое рассмотрение возможных типов входов оказывается не эффективным. Например, в (McDonald, 1993) рассматривалось три вида возможных входов для систем ГЕЯ – числовые данные, структурированные объекты и логические формулы. Более конструктивным является изучение входов, с которыми работают экспериментальные ЛМ системы. Особенность практических систем состоит в том, что они обычно используют в качестве входа представления данных, порожденные другими системами для некоторых практических целей, а не созданные вручную разработчиками системы (Paiva, 1988). В этой области накоплен уже довольно большой опыт. Можно выделить три вида таких входов:

А) БД. Особенность этого типа источника состоит в том, что информация не организована для передачи адресату. Тип текста, который можно построить на основе этой информации, и его структура, должны быть определены извне.

Б) Семантическое представление - представление содержания текста, созданное человеком с помощью системы интерфейсного типа “человек - компьютер”, т.е. такой системы, которая позволяет построить семантическое представление из предлагаемых интерфейсом понятий на основе внутренней речи человека. Этот процесс называется *symbolic authoring*. Правильность полученного семантического представления контролируется системой в процессе

предъявления пользователю очередного фрейма и возможных понятий для заполнения его слотов.

В) Представление знаний на формальном языке, например, SQL, логические языки.

4. Генерация текстов на основе БД

Можно выделить два вида данных БД, которые рассматриваются как источники информации для ЛМ систем ГЕЯ : данные, описывающие некоторые объекты и их признаки, например, БД по товарам, однотипным объектам типа военных кораблей, подержанных автомобилей для продажи и т.п. Другой вид – это поток данных, отражающий состояние одного или группы однотипных объектов в некоторые моменты времени, например, метеорологические замеры (погода), статистические данные по занятости населения (занятость) и др. Эти данные обычно создаются некой лингвистической системой в процессе обработки ею физических или математических данных как вид мониторинга объектов.

4.1. Поток данных - “простой отчет”. Планирование содержания

Содержание текста выбирается из исходной БД. В него попадает только та информация, которая интересует пользователей отчетов. Выбранная информация упорядочивается локально по тематическому принципу, заданному извне, а подача информации в целом соответствует зафиксированному в БД потоку данных. Под типом “простой отчет” мы понимаем отчет, содержащий информацию об одном или некотором множестве однородных объектов с взаимно независимыми параметрами, например, скорость ветра, направление ветра, осадки и т.д. объекта “погода”. К ЛМ системам ГЕЯ, генерирующим простые отчеты, относятся, например:

ANA (Kukich, 1983) выдает биржевой отчет (объем продаж, состояние на момент закрытия, изменение индекса Доу Джонса и т.д.) на основании таблицы данных фондовой биржи за день.

FoG (Goldberg et al., 1994) генерирует метеорологические сводки для кораблей на основании таблицы замеров атмосферных параметров (направление и сила ветра, температура и др.)

PostGraphe (...) - по статистической табличной информации о прибыли компаний синтезирует текстовый отчет и график динамики прибыли.

На рис.1 показано, как исходные понятия (колонка 0) обобщаются и интерпретируются как понятия МПО (колонка 1), а затем собираются в конструкции “объект – значения параметров”, например, *wind 15-20 southwest*, и соединяются предикатами, обобщающими значения параметров в тенденции изменения параметров (колонка 3). В результате фильтрации и резюмирования входных данных, 19 входных значений скорости и направления ветра в системе FoG сводятся к двум событиям МПО. Результатом лексико-грамматической реализации является предложение - один из фрагментов текста: *Wind 15-20 southwest diminishes to wind light southwest*.

0) Исходные данные			1) Понятия ПО		2) Элементы текста
время	Напр. ветра	Скорость ветра	Скорость ветра	Напр. ветра	
7 a.m.	235	17	15-20	southwest	wind 15-20 southwest
9 a.m.	231	21	15-20	southwest	diminish to
...	
9 p.m.	280	12	light	(west)	wind light southwest
10p.m.	307	11	light	(northwest)	
11p.m.	182	8	light	(south)	
12p.m.	246	10	light	(southwest)	

В простых отчетах могут моделироваться простейшие анафорические ссылки в виде замены описательной номинации объекта на указательную - личное местоимение. Ниже приведен фрагмент текста, порожденного системой PostGraphe. В тексте подчеркнуты номинации объекта и анафорические ссылки:

De 1987 à 1988 les profits de B ont augmenté de 160\$ à 165\$. Pendant 1 année ils ont diminué de 25\$. Jusqu'en 1990 ils ont augmenté de 140\$ à 155\$. (цит. по (Paiva, 1998))

Теоретически тематическая структура текста осмысливается как макро-структура текста, своего рода жесткий фрейм со слотами. В (Hasan, 1978) было сформулировано понятие такого фрейма - *generic structure*, состоящего из “стадий” – слотов этого фрейма. Каждая стадия характеризуется единством темы и определенными лексико-грамматическими средствами, соответствующими контексту в рамках данной стадии. Этот подход использовался, в частности, в системе *Komet* (Bateman, Teich, 1995), которая порождала биографические справки о художниках на основании биографической информации из БД.

В некоторых случаях планирование основано на конкретной интеллектуальной операции и может быть реализовано в виде ИИ технологии, например, система SumTime (Spirada et al., 2003), разработанная в Университете Абердина (Шотландия). Данная ИИ технология предназначена для наиболее адекватной передачи пользователю фактической числовой информации об изменении состояния некоторого объекта через определенные промежутки времени. Объекты могут быть различными, в частности, разработаны системы генерации метеорологических сводок, мониторинга газопровода, мониторинга состояния больного. Содержание отчета планируется системой автоматически по числовым табличным данным, при этом система следует постулатам Грайса о хорошо построенном тексте (Grice, 1975). Метеорологическая модель SumTime используется в коммерческом режиме.

4.2. Тексты типа “связный отчет”. Тема-рематическая организация предложения

Связный отчет описывает ситуацию, характеристики которой - различные объекты. Текст создается на основе БД по тому же общему сценарию, что и простой отчет. Различие состоит в том, что дискурс текста представляет собой не просто последовательность тематических блоков или примитивный тематический фрейм типа *generic structure*, а некоторую структуру, образованную семантическими и концептуальными связями – план текста.

Связность моделируется через коммуникативную организацию предложений или динамику фокуса внимания в тексте, т.е. переключение внимания читателя с одного объекта на другой в ходе развертывания текста. Процесс формирования тема-рематической структуры предложений подчиняется определенным правилам. В частности, для управления фокусом внимания в системе ТЕХТ использовались следующие правила, адаптированные из (Sindner, 1979), упорядоченные по предпочтительности:

1. Переместить фокус на объект, упомянутый в предшествующей пропозиции
 2. Сохранить фокус
 3. Вернуться к теме предшествующего обсуждения
 4. Выбрать пропозицию, имеющую наибольшее число имплицитных связей с предшествующей пропозицией.
- Правила формирования тема-рематической структуры предложений при планировании текста обсуждаются также в (Иорданская, 1992). Некоторые системы генерации связных отчетов: *Gossip* (Иорданская, 1992) генерировала отчеты об использовании машинного времени пользователями на основании таблицы данных, в которой указаны идентификатор пользователя, устройство, имена файлов, начало, конец работы с ними, характер работы, время простоя и т.п.

Пример текста из системы *Gossip*: *The system was used for 7 hours 32 minutes 12 seconds. The users of the system ran compilers and editors during this time. The compilers were run six times, for 47% of the cpu time. The editors were run twelve times, for 53% of the cpu time. Two users, Jessie*

and Martin, logged on to the system. Jessie used the system for 63% of the time in use. Martin used the system for 40% of the time in use. (пример из (Paiva, 1998))

PlanDoc (McKeown et al., 1994). *PlanDoc* генерировала отчеты о расширении телефонной сети на основании лог-файла (tracking file), полученного как побочный результат планирования инженерами расширения телефонной сети с помощью специальной программы PLAIN.

Caption Generation System (Mittal et al., 1995) - порождает графики и одновременно текстовые писания этих двумерных графических объектов (графиков).

LFS (Lavoie, Rambow, 1997) генерировала статистические отчеты об изменении параметров занятости населения за месяц на основании БД (занятые, незанятые, уволенные по различным причинам (возраст, пол) и т.п.).

4.3. Генерация описаний и дискурсивные стратегии

Текст, представляющий собой описание некоторого объекта или ситуации, имеет структуру, которая определяется, в первую очередь, свойствами информации самого этого объекта или ситуации. Опыт описаний представлен определенной логикой описаний, закрепленной культурой, которую К. Маккьюин назвала дискурсивной стратегией (Маккьюин, 1989).

Дискурсивные стратегии описания объектов по информации из БД, предложенные Маккьюин, являются реализациями трех коммуникативных целей: Определить, Описать или Сравнить.

Для описания дискурсивных стратегий используются предикативные схемы, составленные в терминах риторических предикатов.

Риторические предикаты характеризуют акты предикирования, которые может использовать говорящий. Они отражают структурное отношение между двумя пропозициями (или фрагментами большего размера) в тексте. Каждому риторическому отношению приписывалось одно из множества значений, таких, как *атрибутив*, *аналогия*, *состав*, *сравнение*, *тема*, *вывод*, *идентификация*, *переименование* и др.).

Предикативная схема - это образец, определяющий, как должен быть построен тот или иной план текста, используя более мелкие схемы или атомарные сообщения и дискурсивные отношения между членами схемы. В системе ТЕХТ в виде схем были представлены четыре различных предикативных модели: *идентификации*, *состава*, *атрибутивности* и *противопоставления*. Работа системы ТЕХТ с использованием предикативных схем выглядела следующим образом: получая на входе коммуникативную цель (*определить*, *описать* или *сравнить*), система выбирала предикативную схему в соответствии с той информацией, которая имеется в БД, и по выбранной схеме устанавливала в общей форме, что должно быть сказано, и в каком порядке.

Особенность планирования содержания в этих типах текстов состоит в том, что оно частично задается извне в виде коммуникативной цели, частично следует свойствам описываемого объекта. Прямое отношение к планированию содержания текстов в терминах дискурсивных стратегий имеют эксперименты по генерации текстов, описывающих пространства, например, интерьер комнаты, в исследованиях по когнитивной лингвистике. В этом случае описание начинается, например, от массивного, центрального объекта, существенными оказываются такие когнитивные свойства как вертикальность, автономность и т.п.

5. Генерация текстов из семантического представления

Кардинально отличается от БД другой источник содержания текста – семантические представления. Этот вид исходных данных создается человеком в режиме интерфейса с компьютером.

Моделирование структуры текста обычно выполняется для рассматриваемых систем в специальной системе планирования содержания, имеющей вид, в частности, графического редактора. Графический редактор последовательно предоставляет пользователю возможность выбирать понятия из определенных в данной семантической среде, следуя заданной в нем стратегии организации текста. Таким образом, графический редактор контролирует правильность получающегося представления, предлагая пользователю для продолжения структуры текста допустимые по структурным и семантическим свойствам понятия МПО. Примерами таких систем являются AGILE (Automatic Generation of Instructions on Languages of the Eastern Europe) – система генерации софтверных руководств к графическим редакторам (международный проект координируемый Институтом исследований по информационным технологиям (ITRI) Университета Брайтона (Великобритания)), и MDA (Multilingual Document Authoring) (Xerox Research Centre Europe) – система, генерирующая различные документы.

Наличие заранее определенных списков понятий, из которых пользователем формируются входные данные, прокламируемая языковая инвариантность и интерлингвистичность представления данных роднит эти системы с технологиями МП, основанными на знаниях - Knowledge-based MT и Interlingua based MT (Nirenburg, 1987). Источниками понятий при создании содержания текста являются онтологии, природа которых варьируется от относительно языково-независимой, как в системе *Patent Claim Expert*, порождающей тексты патентных заявок, до языково-зависимой как онтология Upper-Model в среде KPMML, которая является онтологией семантических типов. Особенность систем данного типа состоит в получении в результате планирования содержания целого связного представления текста, структуру которого образуют дискурсные и предметные отношения между пропозициями или высказываниями.

В отличие от генерации текстов отчетов задача микропланирования в ЛМ системах с семантическим входом состоит, как правило, не в добавлении, а в усечении части исходного семантического представления при переходе к средствам ЕЯ. Переход от единой структуры текста к последовательности предложений, а также реализация некоторых синтаксических конструкций предполагает сокращение фрагментов исходного представления, делая его более лаконичным и естественным. Эти процессы получили название агрегация. При агрегации сокращаются дублирующиеся структуры и понятия.

Содержание может редуцироваться также для ориентации на определенного адресата, например, в системе AGILE есть три режима (стиля) выходного текста: 1) содержание генерируется полностью; 2) редуцируются элементы “методов”, режим ориентирован на хорошо владеющего компьютером пользователя; 3) сохраняются все шаги методов, но редуцируются цели, режим ориентирован на буквальное выполнение действий, “на солдата”.

Пример текста, сгенерированного системой AGILE:

Чтобы нарисовать мультилинию

5. Запустите команду MLINE, выбрав пункт Multiline в палитре Polyline на панели инструментов Draw.
6. На экране появится диалоговое окно Multiline.
7. В строке команд введите “st”, чтобы выбрать стиль.
8. Введите “?”, чтобы показать список стилей.

6. Генерация текстов из представления на формальном языке

Некоторые виды текстов существуют и используются на некотором формальном языке, отличном от языковой семантики. Это языки математики, например, логические языки, языки спецификаций, например, представления запросов к БД (SQL). Особенность “текстов” на этих формальных языках состоит в том, что они коммуникативно организованы, т.е. так же,

как и тексты на ЕЯ, непосредственно предназначены для передачи информации. В качестве входа для системы ГЕЯ такие представления получаются в результате работы определенной лингвистической системы. Примеры систем генерации из формального представления: *Proverb* [...] - текстовое описание логических доказательств и выводов.

Системы перифразирования запроса к базе данных, выраженного на языке реляционной алгебры в ЕЯ:

REMIT (Lowden and Roeck, 1986)

система без названия Минока (Minock, 2003)

QGen (Болдасов, Соколова, 2002)

Для генерации текстов из представления на лингвистическом формальном языке свойственна технология переписывания структуры (structure rewriting).

Литература

1. Болдасов М.В., Соколова Е.Г. Генерация содержания запроса пользователя системой InBASE в виде естественно-языкового высказывания // *Труды Международного семинара Диалог'2002 по компьютерной лингвистике, Том 2, Протвино, Июнь 2002, С. 49-60.*
2. Иорданская Л. (1992) *Коммуникативная структура и ее использование в системе текстовой генерации* // Международный форум по информации и документации. Т. 17, №2. С. 15-28
3. Макьюин К. Дискурсивные стратегии для синтеза текста на естественном языке // *НЗЛ. Вып. XXIV. М., 1989. С. 311-356*
4. Соколова Е.Г., Шаров С.А. (1998). *К многоязыковой генерации руководств пользователя: начальный этап проекта AGILE.* // Труды Международного семинара Диалог'98 по компьютерной лингвистике и ее приложениям (ред. А.С. Нариньяни), Казань. С. 848-859.
5. Соколова Е.Г., Шаров С.А. (2000) Многоязыковая генерация руководств пользователя на славянских языках // *КИИ'2000, Т.1, М.: Изд-во физ-мат. лит-ры С. 409-417.*
6. Bateman, J., Teich, E. (1995) Selective information presentation in an integrated publication system: an application of genre-driven text generation. *Information processing & management, Vol. 31, No. 5, pages 753-767.*
7. Bateman, J. and Zoch, M. (2003) Natural language generation // *The Oxford handbook of computational linguistics* (Ruslan Mitkov ed.) N.Y.: Oxford university press, pp. 284-304
8. Grice, H.P. (1975) *Logic and conversation.* In Cole P. and Morgan J. (Eds), *Syntax and Semantics: Vol 3, Speech acts.* Academic press, New York, pp.43-88
9. Hasan, R. (1978) Text in the Systemic-Functional model // in Dressler (ed.) "Current trends in text linguistics", de Gruyter, Berlin, pp. 228-246.
10. Jokinen, K., Maybury, M., Zock, M. and I. Zukerman (1996) *Gaps and bridges: new directions in planning and natural language generation* // *Proceedings of the ECAI 96 Workshop: Gaps and bridges: new directions in planning and natural language generation.*
11. Kruijff, G.-J., Teich, E., Bateman, J., Kruijff-Korabayova', I., Skoumalova', H., Sharoff, S., Sokolova, L., Hartley, T., Staykova, K. & Hana, J. (2000) *A multilingual system for text generation in three slavic languages.* In *Proceedings of the 18th Conference on Computational Linguistics (COLING 2000)*, Universita't des Saarlandes, Saarbrücken, Germany, pp. 474-480
12. McDonald, D.D. (1997) *Issues in the choice of a source for natural language generation* // *Computational Linguistics, 19(1), 191-197.*
13. Minock, M.J. (2003) A phrasal generator for describing database queries // *Proc. of the 9th EACL workshop on natural language generation, Budapest, Hungary, April 2003.*
14. Lowden, B.G.T. and A.N. Roeck (1986) The REMIT system for paraphrasing relational query expressions into natural language // *Proc. of VLDB (Conference on Very Large Data Bases), Kyoto, Japan, August 1986.*

15. Paiva, D.S. (1998) *A survey of applied natural language generation systems*. Technical report ITRI-98-03, Information Technology Research Institute, University of Brighton, Brighton, UK.
16. Reiter, E. (1995) NLG vs. Templates. In Proc of the Fifth European Workshop on Natural-Language Generation (ENLGW-1995). Leiden, The Netherlands.
17. Reiter, E. and R. Dale (2000) *Building Natural Language Generation Systems*. Cambridge University Press.
18. Sidner, Candace L., 1979. *Towards a Computational Theory of Definite Anaphora Comprehension in English*. Ph.D. thesis, AI Laboratory / MIT, Cambridge, MA. Also available as Technical Report No. AI-TR-537 (Spirada et al., 2003),
19. The John Bateman and Michael Zock's list of Natural Language Generation Systems: <http://purl.org/net/nlg-list>
20. SIGGEN (Special Interest Group on GENeration) – специальная группа по интересам в рамках Ассоциации по компьютерной лингвистике (ACL), которая доступна через сайт ACL: <http://www.cs.columbia.edu/~acl/>
21. KPML (Komet Project MultiLinual) – Тактический генератор - специальная среда для генерации текстов на различных языках. Имеются ресурсы для английского, французского, немецкого, голландского, японского, чешского, русского и др.: <http://purl.org/net/kpml>
22. CoGenTex – фирма, специализирующаяся в области ГЕЯ (США-Канада), специализировалась на генерации отчетов. Основная задача – использование научных разработок для создания практических систем ГЕЯ: <http://www.cogentex.com/research/modex/index.shtml>
23. ITRI – Институт исследования информационных технологий Университета Брайтона (Великобритания), который занимается коммуникацией человека с компьютером на естественном языке. ITRI является организатором конференции INLG04, которая состоится в Нью Форест (Великобритания) в июле 2004 г. <http://www.itri.brighton.ac.uk/>