

НАУЧНАЯ ЖИЗНЬ

КОНФЕРЕНЦИИ, ПОСВЯЩЕННЫЕ КОМПЬЮТЕРНОЙ (ИНЖЕНЕРНОЙ) ЛИНГВИСТИКЕ

Стремительный рост научно-технических, торговых и культурных связей между различными регионами мира и особенно экономическая экспансия "тигров Азии" (Японии, Южной Кореи, Китая и государств Юго-Восточной Азии) стимулировали интенсивное развертывание практических и георегических работ в сфере автоматической переработки деловых и научно-технических текстов — в первую очередь в области машинного перевода (МП), автоматического реферирования и аннотирования, а также в лингвистических аспектах искусственного интеллекта. Новые результаты в этих областях знаний обсуждались на многочисленных научных конференциях, имевших место в последние два года (1990—1992) в разных странах. Среди них наиболее интересными для языковедов являлись следующие форумы.

1. Две конференции, организованные международным комитетом по компьютерной лингвистике (International Committee on Computational Linguistics). Первая из них (COLING-90) проходила 20—25 авг. 1990 г. в Хельсинки, а вторая (COLING-92) состоялась 20—28 июля в г. Нант (Франция).

2. Две объединенные конференции Ассоциаций "Компьютеризация литературоведческих и лингвистических исследований" (Association for Literary and Linguistic Computing — ALLC) и "Компьютеры и гуманитарные науки" (Association for Computers and the Humanities — ACH). Первая из них (ACH-ALLC-91) проходила 17—21 марта 1991 г. в Аризонском университете (г. Темпе, США); а вторая (ALLC-ACH-92) имела место 5—9 апреля 1992 г. в Оксфорде (Великобритания).

3. Международная конференция по текущим вопросам компьютерной лингвистики (ICCL), организованная Малайзийским университетом. Она проходила 10—14 июня 1991 г. в г. Пенанг (Малайзия).

4. Международная конференция по терминологии, стандартизации и обмену технологиями (ICISI-91), состоявшаяся 2—6 июня 1991 г. в Пекине. Она была организована Китайским национальным техническим комитетом по терминологии и стандартизации.

5. 3-я конференция по прикладной автоматической переработке текста (ANLP-92). Эта конференция, организованная Трентийским институтом культуры и Институтом научно-технологических исследований Италии, проходила с 31 марта по 3 апреля 1992 г. в г. Тренто (Италия).

6. Международная конференция "Инженерная лингвистика и оптимизация преподавания языков" (ИЛОИЯ), проходившая 28—30 мая 1992 г. в Самарканде (организована Самаркандским университетом им. А. Навои).

7. 4-й Международный коллоквиум по теоретическим аспектам машинного перевода "Эмпирия и рационализм" (TMI-92) был созван Канадским исследовательским центром по информатизации труда и состоялся 25—27 июня 1992 г. в Монреале.

8. 8-й симпозиум "Лингвистические проблемы искусственного интеллекта" (ЛПИИ), организованный Научно-техническим обществом радиотехники, электроники и связи им. А.С. Попова и Российским педагогическим университетом им. А.И. Герцена (27 окт. 1992 г., С.-Петербург).

Перечень публикаций по названным конференциям дан в конце обзора I—II.

На конференциях обсуждался широкий круг кибернетических, программистских, психологических и социологических вопросов, многие из которых выходили за пределы теории языка и инженерной лингвистики (ИЛ), рассматриваемой как теория и практика компьютерного решения лингвистических задач. Для языковеда наибольший интерес представляют следу-

ющие четыре направления, охватившие основную массу докладов и выступлений.

I. Когнитивные и лингвистические аспекты искусственного интеллекта (ИИ). Здесь в центре обсуждения оказались проблемы

— инженерии энциклопедических знаний, лингвистической (лексико-грамматической) компетенции, их организации и хранения в памяти компьютера, — ср. доклады Х.С. Чан (Ю. Корея), И.А. Болынакова (Москва) — COLING-90; Е.В. Падучевой, Е.В. Рахилиной, М.В. Филипенко (Москва), Ф. Русло и Б. Миго-Ансу (Франция), В.Г. Андреэна, В.А. Квятковского и др. (С.-Петербург), Б. Блазер, А. Сторрер и У. Швалль (ФРГ), Д. Фарвелла, Л. Гутри, Й. Уилкса (США) — COLING-92; М. Дигэн (Великобритания), Ж. Верониса и Н. Ид (Франция), Л.Г. Хоппертона и Н.Ж. Прюбер (Канада) — ALLC-ACH-92 и TMI-92; Л. Карлсона и С. Ниренберга (США) — ANLP-92; Т.А. Аполлонской и Л.Н. Беляевой (С.-Петербург) — ИЛОПЯ, ЛПИИ;

— семантико-синтаксического и прагматического анализа и синтеза текста, включая вопросы "наивной" семантики и немонотонного (нестрого логического) вывода (подробнее см. [12]); см. проект Дж.К. Мьерса (Япония) — COLING-90; сообщения Ф. Сегона и К. Енсен (США), В. Задорожного (США) — COLING-92; А. Лавелли, Б. Магинни, К. Страппараца (Италия) — ANLP-92; В.В. Богданова и Э.И. Соболевой (С.-Петербург) — ЛПИИ;

— извлечение из текста его основного содержания (т.е. автоматическое реферирование, аннотирование или индексирование текста) — см. доклады Р.Дж. Куна (США) — COLING-90; Б. Цой и др. (США), Л. Шмитт, Ж. Руйот, Э. Оливетан (Франция) — COLING-92; Дэ.С. Чанга и др. (Китай) — ANLP-92; С.П. Грицай, В.Р. Нымм и др. (С.-Петербург) — ИЛОПЯ.

II Технологии формального грамматического анализа. Здесь внимание грамматистов привлекают доклады, посвященные использованию в работающих системах МП таких современных и плохо известных в отечественном языкознании концепций, как:

неохомскийская грамматика управления и связывания (Government and Binding Grammar) — см. сообщения Х.Х. Ченя (Тайвань), А. Джорджи, Ф. Пьянези и Дж. Сатта (Италия) — COLING-92;

— грамматика присоединяемых деревьев (Tree Adjoining Grammar) — см. сооб-

щения К. Харбуш (ФРГ) — COLING-90; К. Вияв-Шанкер, И. Шабеса, П. Параубек (США) — COLING-92 и ANLP-92;

— грамматика фразовых структур (Phrase Structure Grammar) — см. сообщения Л. Аренберга (Швеция), Ф. Бланша и Ж.-И. Морена (Канада) — COLING-92; доклад Р. Кёлера (ФРГ), А. Нахимовского (США) и др. — ИЛОПЯ.

III. Перспективы создания интеллектуального лингвистического автомата (ЛА), так называемого *homunculus loquens*, способного осуществлять разные виды анализа и синтеза (текста). Эта проблема детально обсуждалась на нескольких конференциях. Начало было положено в ходе состоявшейся во время Хельсинкского совещания COLING-90 дискуссии по вопросам соотношения естественного языка (ЕЯ) и языка компьютера (ЯК). Выступавшие во время дискуссии Х. Карлгрен (Швеция), В. Хаан (ФРГ), Л. Гиришман, Й. Уилкс (США), Кр. Буате (Франция) подчеркивали, что создание интеллектуального ЛА связано с ослаблением барьера, отделяющего "открытый" нечетко-множественный ЕЯ от ЯК, характеризующегося жесткой закрытой структурой (ср. в этом плане доклад Т.А. Аполлонской и Л.Н. Беляевой, прочитанный на совещании ЛПИИ). Предлагалось использовать вероятностно-статистические приемы для ослабления этого барьера. Принявший участие в Хельсинкской дискуссии Р.Г. Пинотровский (С.-Петербург) напомнил, что вероятностно-статистические приемы, широко используемые уже с конца 60-х годов в советской и российской инженерной лингвистике при построении промышленных систем МП, а также систем реферирования и индексирования текста, не должны ограничиваться примитивной лексико-грамматической статистикой. Они дают ощутимый эффект, во-первых, при учете распределения лексико-грамматических единиц в тексте, а во-вторых — в сочетании с данными об информационной структуре текста и образующих его сегментов (ср. [13—15]).

Идеи Хельсинкской дискуссии нашли свое конкретное развитие на последующих форумах. Так, Дэ. Цулзин (Великобритания — Япония) в своем программном докладе "Современное состояние систем МП" — ANLP-92 подчеркнул, что в современной парадигме машинно-переводческих стратегий (особенно при создании коммерческих систем) вероятностно-статистический перевод (Statistic-based MT) займет, очевидно, центральное место вмес-

ге с такими новыми подходами, как перевод по готовым образцам (Example based MT) — см. доклады Д. Джоунза (Великобритания), Х. Маруяма и Х. Ватанабе (Япония) — TMI-92 и перевод по аналогиям (Analogy-based MT). Информация о конкретном применении вероятностно-статистической методики при построении систем МП содержалась в докладах Р. Резника (США), И. Эрран и др. (Франция), Л. Иорданской и др. (Канада) — COLING-92; П. Брауна и др. (США), Х. Лемана и Н. Отта (ФРГ) — TMI-92; Н.М. Алексеева (С.-Петербург) — ИЛОПЯ.

IV. Применение инженерно-лингвистических технологий и лексикограмматических формализмов к иноструктурным (в первую очередь — восточным языкам). Здесь прежде всего обращают на себя внимание следующие темы:

— использование традиционных приемов грамматического анализа (парсинга) слова и предложения в языках с внутренней флексией (арабский, иврит), в агглютинативных языках, использующих грамматику порядков, а также в изолирующих языках, — см. доклады А. Эллаггара (Египет), Г. Сатоши (Япония), Х. Эль-Шишния (Египет), К. Коскенниemi (Финляндия) — COLING-90; С. Фонт и Р.К. Бервик (США), Х.Р. Юсуф (Индонезия);

— COLING-92: М. Nagao (Япония) — ICCIC; А. Окимура, К. Мураки, К. Ямавана (Япония) — TMI-92;

— организация лингвистических баз данных и автоматических словарей, а также построение лингвистического автомата, ориентированного на восточные языки, — см. доклады Ч.Ж. Хуанг, К. Чень (Тайвань) — COLING-92; Р.Г. Пиотровского, Р.С. Минвалеева, В.Г. Шумовского и др. (С.-Петербург) — ACH-ALCC-91 и ICTST-91; Х. Ясунага (Япония) — ALCC-ACH-92, А.И. Чапли и др. (Израиль) — ИЛОПЯ.

В ходе конференций зарубежные фирмы и коллективы демонстрировали экспериментальные и промышленные системы машинного перевода. Среди представленных на этих совещаниях научных коллективов стран СНГ только группа "Статистика речи" сумела показать реально работающие системы англо- и франко-русского МП (COLING-90, ИЛОПЯ, ЛПИИ).

Особое внимание во время конференций COLING-92 (Нант), ICCIC (Пенанг), ALLC-ACH-92 (Оксфорд), ANLP-92 (Тренто), ИЛОПЯ (Самарканд) и TMI-92 (Мон-

реаль) было уделено вопросам компьютерной поддержки преподавания языков, а также машинных текстовых фондов, предназначенных для лексико-грамматических, семантико-синтаксических и фонологических исследований литературного языка и его истории, в том числе для изучения языка писателей.

Итак, лингвистическая инженерия, прощая шаг за шагом во все новые области филологии, не только расширяет экспериментальные возможности этой науки, но и стимулирует возникновение в ней новых теоретических проблем.

СПИСОК ЛИТЕРАТУРЫ

1. Инженерная лингвистика и оптимизация преподавания языков. Республиканская конференция, 28—30 мая 1992 г.: Тез. докл. Самарканд, 1992.
2. Лингвистические проблемы искусственного интеллекта: Материалы VII симпозиума, 27 октября 1992 г. СПб., 1992.
3. COLING-90: Papers presented to 13-th International conference on computational linguistics on the occasion of the 25-th anniversary of Helsinki University / Ed. Karlgren H. Helsinki, 1990.
4. COLING-90: 13-th International conference on computational linguistics. August 20—25, 1990: Programme and Exhibition catalogue. Helsinki, 1990.
5. Fourteenth International conference on computational linguistics. COLING-92: Programme, 20—28 July 1992. Nantes. 1992.
6. 1991 Joint conference of the ACH and the ALLC. March 17—21 1991: Programme. Tempe (Ariz), Arizona State University, 1991.
7. 1992 Joint conference ALLC-ACH-92, 5—9 April 1992: Programme. Oxford, 1992.
8. Proceedings of the International conference on current issues in computational linguistics. 12—14 June 1991. Penang (Malaysia), 1991.
9. International conference on terminology, Standardization and technology transfer: Proc. 1991-02-02/06. Beijing, 1991.
10. ANLP-92: 3-rd Conference on applied natural language processing. Trento, Italy, 1—3 April 1992: Programme. Trento, 1992.
11. Quatrième Colloque international sur les aspects théoriques. TMI-1992: Programme préliminaire. Méthodes empiristes versus méthodes rationalistes en TA. 25 au 27 juin 1992. Montréal, 1992.
12. Dahlgren K. Naïve semantics for natural language understanding. Boston; Dordrecht; London, 1988.
13. Статистика речи. Л., 1968.
14. Статистика речи и автоматический анализ текста. Л., 1971, 1973, 1974, 1978, 1980.
15. Piotrowski R., Leschriin M., Lukjanenkov K. Introduction of elements of mathematics to linguistics. Bochum, 1990.

Пиотровский Р.Г. (С.-Петербург)