

**БІЛОЦЕРКІВСЬКИЙ ДЕРЖАВНИЙ АГРАРНИЙ УНІВЕРСИТЕТ
ВІДКРИТИЙ МІЖНАРОДНИЙ УНІВЕРСИТЕТ
РОЗВИТКУ ЛЮДИНИ "УКРАЇНА"
(Білоцерківське міське територіальне представництво)**

**Європейські орієнтири освіти
України в умовах глобалізації
(матеріали науково-практичної конференції,
4-6 червня 2004 р., м. Біла Церква)**

**Біла Церква
2004**

О.П.ЦВИД-ГРОМ, канд.філол.н, завідувач кафедру документознавства та інформаційної діяльності Білоцерківської філії ВМУРоЛ “Україна”

ДО ПРОБЛЕМИ ЛІНГВІСТИЧНОГО ЗАБЕЗПЕЧЕННЯ ІНФОРМАЦІЙНИХ ПРОЦЕСІВ У СИСТЕМІ ДОКУМЕНТНО- КОМУНІКАЦІЙНОГО ЦИКЛУ

Комп'ютеризація суспільства у всіх сферах його діяльності, ліквідація комп'ютерної неграмотності, всебічне і глибоке оволодіння обчислювальною технікою стали важливими факторами прискорення науково-технічного прогресу. Водночас, перед науковцями постало ряд завдань лінгвістичного характеру, що можуть бути успішно вирішені лише при розумному поєднанні інтелектуальних здібностей людини і можливостей сучасної електронно-обчислювальної техніки.

У зв'язку з застосуванням ЕОМ лінгвістичні методи, що використовуються для вирішення задач НТІ, при потребі, трансформуються і набувають специфічних рис комп'ютерної лінгвістики - науки, що вивчає мову і мовлення з метою їхнього застосування в системах автоматизованої обробки інформації.

Проблеми комп'ютерної лінгвістики вивчає Пещак М.М., Піотровський Р.Г., Белоногов Г.Г., Кузнецов Б.А., Блехман М.С. Арапов М.В., Бакулов А.Д., Чорний А.І. та інші. Нагальними питаннями у системі лінгвістичного забезпечення інформаційних процесів є: складання і лінгвістична обробка машинних словників, виявлення і виправлення помилок у текстах під час їхнього введення в ЕОМ, індексування документів і інформаційних запитів, класифікація і реферування документів, лінгвістичне забезпечення процесів пошуку інформації в одномовних і багатомовних базах даних, машинний переклад текстів з одних природних мов на інші і т.д. Важливою науковою проблемою залишається моделювання процесів розуміння змісту текстів (перехід від тексту до формалізованого представлення його змісту) і синтезу мови (перехід від формалізованого представлення змісту до текстів природною мовою).

Лінгвістичні засоби, створювані і застосовувані в комп'ютерній лінгвістиці, умовно поділяють на дві частини: декларативну і процедурну [2,37]. До декларативної частини відносяться словники одиниць мови і мовлення, тексти і різного роду граматичні таблиці, до процедурного - засоби маніпулювання ними. Важливою методологічною проблемою прикладної комп'ютерної лінгвістики є правильна оцінка необхідного співвідношення між декларативним і процедурним компонентами системи автоматичної обробки текстової інформації. Чому віддати перевагу - обчислювальним процедурам, що спираються на відносно невеликі словникові системи з багатою граматичною і семантичною інформацією, чи декларативному компоненту при відносно простих процедурних засобах? На думку вчених (Г.Г.Белоногова, Ю.Г.Зеленкова, Б.А.Кузнецова та інших) другий шлях переважає. Він швидше приведе до вирішення практичних завдань, оскільки менше зустрічатиметься перешкод і можна буде в більших масштабах використовувати ЕОМ для автоматизації досліджень і розробок [3,57].

Перспективним є такий шлях розвитку комп'ютерної лінгвістики, коли основні зусилля будуть спрямовані на створення могутніх словників одиниць мови і мовлення, вивчення їхньої семантико-синтаксичної структури і розробку базових процедур морфологічного, семантико-синтаксичного і концептуального аналізу і синтезу текстів.

Машинний словник невід'ємна складова будь-якої системи автоматичної обробки текстової інформації. Вони можуть бути моно- чи політематичними і являти собою словники слів і словосполучень, що виражають стійкі науково-технічні поняття. Слова і словосполучення повинні супроводжуватися граматичною і семантичною інформацією. Під час складання подібних словників необхідно прагнути до того, щоб вони максимально відбивали лексичний склад науково-технічних текстів.

Як уже зазначалося, у системі НТІ є необхідність вирішення задачі автоматизованого виявлення і виправлення помилок у текстах під час їхнього

введенні в ЕОМ. Ця комплексна задача може бути умовно розчленована на три задачі - орфографічний, синтаксичний і семантичний контроль текстів [5,20]. Перша з них реалізується процедурою морфологічного аналізу з використанням досить могутнього еталонного машинного словника основ слів. При цьому, якщо основи слів, що піддаються морфологічному аналізу тексту, і ототожнюються з основами еталонного словника, то вони вважаються правильними, якщо ні, то мікроконтекст із даними словами видається на перегляд людині.

Синтаксичний контроль текстів з метою виявлення в них помилок значно складніший орфографічного. По-перше, тому, що він включає й елементи орфографічного контролю як обов'язковий компонент, а, по-друге, тому, що проблема синтаксичного аналізу неформалізованих текстів у повному обсязі ще не вирішена. Проте, частковий синтаксичний контроль текстів цілком можливий. Тут можна йти двома шляхами: або скласти досить ґрунтовні машинні словники еталонних синтаксичних структур і порівнювати з ними синтаксичні структури аналізованого тексту, або розробляти складну систему правил перевірки, граматичної погодженості елементів тексту. Перший шлях, на нашу думку, є ефективнішим, хоча, звичайно, він не виключає можливості застосування елементів другого шляху. Синтаксична структура текстів повинна описуватися в рамках граматичних класів слів (точніше, у вигляді послідовностей наборів граматичної інформації до слів).

Семантичний контроль текстів з метою виявлення в ньому значеннєвих помилок варто віднести до класу задач штучного інтелекту. У повному обсязі вона може бути вирішена тільки на основі моделювання процесів людського мислення. Для цього, очевидно, слід створювати могутні енциклопедичні бази знань і програмні засоби маніпулювання ними.

Ми вже згадували про труднощі машинного перекладу текстів з одних природних мов на інші. Це і зрозуміло: переклад - складний творчий процес, і його адекватне моделювання на ЕОМ буде можливим лише тоді, коли будуть вивчені основні закономірності цього процесу. У розв'язанні багатьох питань машинного перекладу значних результатів досягли Кулагіна О.С., Падучева О.В., Андрєєв М.Д. та інші. Машинний переклад завжди був у центрі уваги у своїй, майже, 65-річній історії. Переклад людиною з однієї мови на іншу відбувається за допомогою сприйняття і розуміння вихідного тексту і подальшої передачі його змісту засобами вихідної мови. При цьому перекладаються не самі слова і словосполучення, а понятійні образи, що виникають у свідомості перекладача під їхнім впливом [4,33]. Оскільки, поки що немає можливості моделювати роботу людини-перекладача стовідсотково, то потрібно хоча б прагнути оперувати тими одиницями мови і мовлення, що дозволяють найточніше передавати зміст тексту, написаного однією мовою засобами іншої мови. Такими одиницями, на наш погляд, є, насамперед, стійкі

усталені вирази, звороти і термінологічні словосполучення, вже потім - окремі слова.

Якщо сьогодні повний автоматичний високоякісний науково технічний переклад практично неможливий, то машино-людино-машинний переклад цілком реальний. Для цього ЕОМ повинна виступати, насамперед, у ролі бази знань, що містить перекладні відповідності, які найчастіше зустрічаються, - фрази, термінологічні словосполучення і окремі слова, а також розташовувати їх могутньою програмною системою для морфологічного і синтаксичного аналізу і синтезу текстів та їхнього редагування за вказівками людини.

ЕОМ сучасний, могутній технічний засіб, що забезпечує різнобічну інтенсифікацію і оптимізацію освітніх процесів. Уміле використання нових електронних технологій має виключне значення у вирішенні практичних проблем економічного і соціального розвитку. Багато актуальних питань теорії і практики лінгвістичних систем (підвищення якості машинної переробки мовної інформації, подальший розвиток сфер автоматизації процесів міжмовної комунікації, надійний аналіз і синтез текстів тощо) чекають свого розв'язання.

ЛІТЕРАТУРА

1. Автоматическое индексирование и реферирование документов. - М., 1983.
2. Бакулов А. Д., Черный А. И. Лингвистическое обеспечение автоматизированных банков данных. - М., 1981.
3. Белоногов Г. Г., Кузнецов Б. А., Новоселов А. П. Автоматизированная обработка научно-технической информации. Лингвистические аспекты. - М., 1984.
4. Хархалис Р. Компьютерный перевод иностранных текстов. К., 1998.
5. Использование ЭВМ в лингвистических исследованиях. К., 1990. .