

МОВЛЕННЕВА ДІЯЛЬНІСТЬ ТА ЇЇ МОДЕЛЮВАННЯ

Інформаційні природно-мовні технології як моделі мовленнєвої діяльності людини

Мовленнєва поведінка людини постає однією з найскладніших форм інтелектуальної діяльності, яка реалізується в різноманітних проявах. Сучасні інформаційні технології (ІТ) на базі потужної обчислювальної техніки та створених на сьогоднішній день програмних продуктах все ефективніше намагаються моделювати певні аспекти мовленнєвої діяльності (МД) людини.

Можемо вже окреслити цілий кластер інформаційних технологій (далі *інформаційні природно-мовні технології - ІПМТ*), що орієнтовані на моделювання окремих проявів мовленнєвої діяльності людини. Перш за все це будуть: пошук в INTERNET та всі пошукові системи, експертні системи, синтез/аналіз мовленнєвої інформації, синтез/аналіз текстової інформації, автоматичний переклад, природно-мовний інтерфейс, природно-мовні бази знань, системи накопичення знань тощо. Всі перелічені сучасні технології намагаються моделювати лише окремі прояви мовленнєвої поведінки людини – відповідь на поставлені запитання, накопичення та використання знань з певної предметної сфери, генерація та розуміння мовленнєвої або текстової інформації, володіння різними мовами (переклад), глибоке розуміння мовної організації, поступове накопичення знань.

Проте, незважаючи на потужні ресурси сучасної обчислювальної техніки, фінансового забезпечення та людського потенціалу, що вкладаються в ці технології, результати ще й досі не відповідають реаліям мовленнєвої діяльності людини:

- інформаційний пошук, практично, захлинається в океані інформації (великий обсяг нерелевантної інформації отримуємо на запит);
- експертні системи дають непогану узагальнену картину проблеми, яка в конкретному випадку часто спрацьовує неадекватно;
- синтез мовного повідомлення (у чистому вигляді – як результат роботи нашого мозку) поки що відсутній, аналіз – пов'язаний все ще з багатьма проблемами;
- синтез мовлення як моделювання мовленнєвого тракту людини – практично єдиний приклад інформаційних технологій, що вийшов на гарний комерційний рівень – отже, задовольняє вимогам користувачів, і ми вже сьогодні можемо придбати синтезатор, практично для довільної мови;

- аналіз мовленнєвого повідомлення (особливо безперервного) все ще стикається з багатьма проблемами і не завжди задовольняє користувача;
- автоматичний переклад - головним чином реалізований в людиномашинному варіанті і кваліфіковані перекладачі не користуються ним;
- природно-мовний (ПМ) інтерфейс, у загальному випадку, використовуватися може лише для оформлення коротеньких послань електронної пошти і не більше, де допустимі невеличкі похибки стилістичного та семантичного плану (числова інформація поки що сприймається неадекватно);
- природно-мовні бази знань (на відміну від існуючих баз даних), практично, не мають чіткої стратегії розвитку і не виходять поки що за межі фіксації найпростіших відношень онтології; накопичення знань в пошукових системах ведеться в ручному варіанті.

Отже можемо констатувати: майже за 50 – 60 років свого розвитку, моделювання мовленнєвої діяльності не досягло бажаних результатів, і сучасні ПМ-технології все ще неадекватно відтворюють реалії мовленнєвої поведінки. В чому ж тут справа та які проблеми? Відповідь мабуть одна – мовленнєва діяльність, з одного боку, це одна з найскладніших форм інтелектуальної діяльності людини, для розуміння якої потрібно інтегрувати сучасні дослідження у багатьох помежованих сферах (лінгвістиці, психології, філософії, нейрофізіології, кібернетиці, когнітології, інформаційних технологіях тощо), а з іншого – ми й досі ще недостатньо знаємо сам об'єкт досліджень – мовленнєву поведінку людини, що реалізується нашим мозком.

Тут у весь зріст постає стара філософська проблема: чи може мисляча субстанція сама зрозуміти, як вона мислить, як працює та спілкується з подібними собі суб'єктами? Це питання непогано проаналізоване в монографії “Від думки до знання” [1]. Автор не може дати конкретної відповіді, проте досвід засвідчує, що все більш інтенсивне накопичення знань стосовно складних форм існування та функціонування матерії (і живої, особливо) дозволяє нам створювати все більш адекватні моделі цього явища не лише віртуальні, але й формальні – електронні.

Тож спробуємо трохи по-іншому поглянути на проблеми дослідження мовленнєвої діяльності людини та інтегрувати сучасні досягнення багатьох помежованих напрямів, що тим чи іншим чином пов'язані з аналізом окремих проявів МД. Можливо це допоможе нам більш чітко зрозуміти сучасний стан інформаційних ПМ-технологій, проблеми моделювання та напряму виходу на новий щабель як дослідження мовленнєвої діяльності, так і моделювання окремих її аспектів в ІТ.

Індивідуальна мовна система

Мовленнєва діяльність кожної людини згідно з поглядами Л.В. Щерби [2] актуалізується її індивідуальною мовною системою (ІМС), яка за

визнанням багатьох кваліфікованих фахівців - психологів, філософів і, звичайно ж, лінгвістів – формується з двох частин: перша – охоплює всі знання про мовну організацію (на свідомому, або більше - підсвідомому рівнях), тоді як друга – практично пов'язана з актуалізацією моделі зовнішнього середовища, в якому живе людина. Першу частину, що зумовлює мовну компетенцію індивіда, позначимо як лінгвістичний процесор (ЛП), тоді як другу – вважатимемо за базу знань (БЗ), де концентруються всі накопичені індивідом знання на поточний момент стосовно життєвого досвіду та зовнішнього середовища. У загальному випадку, ЛП охоплює всі існуючі граматичні побудови (класична лінгвістика) та словники (БД - словники), тоді як в БЗ формується модель світу, в якому виховувався та проживає індивід (звісно, модель середовища охоплює власне і модель індивіда, яка представляє його ЕГО і позначається як “Я”). Знову ж таки, посилаючись на Л.В. Щербу, ІМС може функціонувати в режимах синтезу або аналізу повідомлення; у випадку аналізу -лінгвістичний фільтр визначає семантико-синтаксичні прикмети повідомлення, спираючись при цьому як на БД-словники і граматику, так і на знання із БЗ, тоді як синтез повідомлення йде у зворотному напрямку, коли думка, що сформована поточним оточенням та інформаційним станом людини, проходить процес граматизації (формування структури повідомлення) з урахуванням всіх стандартів ЛП. Практично, обидві складові ЛП та БЗ завжди знаходяться у діалектичній єдності і працюють одна на одну: БЗ допомагає ЛП на стадії аналізу, а результат аналізу, в свою чергу, поповнює або корегує наповнення БЗ. У загальному випадку, враховуючи, що головна функція мови – комунікативна, наповнення складових ЛП та БЗ окремих ІМС повинні бути більш-менш ідентичними; особливо це стосується загальних рис ЛП, тоді як наповнення БЗ найчастіше визначається життєвим досвідом конкретного індивіда на поточний момент часу, хоча якась певна частка знань багатьох індивідів повинна збігатися, щоб гарантувати більш-менш ідентичне сприйняття комунікантами довільного повідомлення певної життєвої орієнтації. У випадку розбіжності відповідних фрагментів БЗ обох індивідів в процесі комунікації можливе уточнення окремих складових через постановку запитання на кшталт: “А що ти розумієш під цим?”.

Отже, можемо констатувати, що процес комунікації на мовному рівні актуалізується індивідуальною мовною системою в режимах синтезу або аналізу повідомлення з урахуванням діалектичної єдності лінгвістичного процесора як сукупного знання індивіда стосовно мовної організації та бази знань, де зберігається весь життєвий досвід про світ та життєві колізії.

Сучасний стан та перспективи інформаційних природно-мовних технологій

Тепер вже з позицій розуміння структурно-функціональної організації ІМС можемо оцінити сучасний стан та перспективи інформаційних технологій, що намагаються моделювати ті чи інші прояви мовленнєвої

діяльності людини. Коли лінгвістична спільнота визнає, що мовленнєва діяльність людини актуалізується ІМС, то фахівці, що намагаються дослідити та моделювати найважливіші прояви функціонування ІМС, авторитетно заявляють (зокрема Мельников Г.П. [3]), що електронна модель ІМС на структурному рівні також повинна мати дві складові M_1 як модель ЛПІ та M_2 як модель БЗ, що замикаються через потужний інтерфейс, який забезпечує взаємодію цих складових.

Який же сучасний стан справ у сфері формування ПІМТ, що моделюють певні прояви мовленнєвої діяльності людини? Можемо відповідально констатувати, що другою складовою ІМС – моделі M_2 як електронної оболонки під природно-мовну базу знань на сьогоднішній день, практично, не існує. Сучасні БД не пристосовані для ефективної роботи з довільною природно-мовною (ПМ) інформацією, і лише останнім часом з'являються перші спроби опанування сферою БЗ через онтології в певних прикладних напрямках (див., наприклад Гаврилова Т.А., Хорошевський В.Ф. [4]). Проте, поки що, це розробки на рівні найвищого узагальнення наших знань про зовнішнє середовище, які дещо нагадують дитячу гру, коли у відповідь на випадково озвучене слово синтезується реакція стосовно належності його змісту до категорій живе/неживе, істивне/неістивне тощо. З таких категорійних оцінок надалі формується семантична мережа для певної предметної сфери. Такий підхід поки що зовсім нерелевантний намаганням формувати знання на рівні опису множини окремих ситуацій, що у сукупності своїй адекватно відтворюють реалії світу або конкретної предметної сфери.

А що можемо сказати відносно першої складової ІМС – моделі нашої мовної компетенції у якості ЛПІ? У загальному випадку, наші знання стосовно мовної організації визначаються класичною лінгвістикою (усіма її розділами), і практично всі сучасні розробки ЛПІ базуються на цих засадах. Яка ж оцінка сучасного етапу розвитку класичної лінгвістики? А загальна думка лінгвістичної спільноти на сьогоднішній день така, що в граматичній науці відсутня системність, і в ній поки що більше виключень ніж правил - це позиція, яка озвучена багатьма авторитетними фахівцями.

Проте сучасний етап розвитку класичної лінгвістики вже не задовольняє потребам формування сучасних інформаційних технологій, орієнтованих на моделювання мовленнєвої діяльності людини. На жаль можемо констатувати, що знання про мовну організацію, які накопичувалися протягом останніх 350 років (з моменту появи першого кваліфікованого дослідження – Граматики Пор-Рояля) поки що багато в чому не адекватні реаліям мовної організації. Найкраще тут звернутися до визнаних авторитетів.

- Уїтні, Ф.де Соссюр вважали мову сукупністю різнопланових одиниць, між якими нема чіткого логічного зв'язку; ще більш критична оцінка – мова це звалище (хаос), де важко розібратися [5].

- Піотровський Р.Г. вважав сучасну граматику “чисто спекулятивною наукою”, побудованою на коротеньких вибірках, що не дозволяє вийти на достойний рівень узагальнення [6].
- Л. Астахова визначає, по-перше, що сучасна лінгвістична спільнота давно вже визріла до заперечення усталеної дихотомії “просте / складне речення” а, по-друге, граматична наука до цього часу все ще не визначила чітко власне сам об’єкт синтаксичних досліджень [7]. Останнє, як і перше – дуже важливі претензії до сучасної синтаксичної науки. Доречно тут нагадати, що на даний момент в лінгвістиці існує близько трьохсот (300!) визначень простого речення, що теж наводить на певні роздуми.
- Граматика -70 констатує, що всі досягнення класичної лінгвістики стосовно мовної організації, які презентовані збірником за десятиріччя, не можуть розглядатися з позицій єдиної системи [8]; тобто в наших знаннях відсутня системність.
- Від себе можемо доповнити, що ще й до цього часу не вирішена однозначно проблема словосполучення.

Враховуючи вищесказане можемо зробити підсумок: всі сучасні проекти лінгвістичного процесора, які створюються або вже створені, лише дуже наближено моделюють особливості мовленнєвої організації людини і, практично, не здатні працювати з довільним мовним матеріалом. Висновок з цього напрошується не зовсім втішний: на основі сучасної класичної лінгвістики ми не в змозі створити більш-менш адекватну модель ІМС, бо практично відсутня модель середовища в БЗ (M_2), а лінгвістичний процесор (M_1), створений на засадах класичної лінгвістики, не може працювати з довільними (не класичними) структурами. Тут абсолютно доречно навести зауваження Золотової Г.А. стосовно людських можливостей пізнання складних залежностей нашого існування: процес пізнання таких явищ реалізується шляхом побудови все більш адекватних моделей, що знімають певні суперечності попередніх [9]. Отже, перед нами дилема – або залишатися в межах усталеної парадигми класичної лінгвістики, яка не в змозі адекватно відтворювати найважливіші функції мовленнєвої діяльності людини, або критично переглянути її основні постулати на основі інтеграції сучасних досліджень багатьох помежованих напрямів, пов’язаних з аналізом мовленнєвої діяльності людини.

Тож приймаємо останній шлях аналізу мовної поведінки людини, бо стає все більш очевидним, що досліджуючи одну з найскладніших форм інтелектуальної діяльності лише за кінцевим продуктом цієї діяльності – мовним матеріалом, не враховуючи при цьому весь складний ланцюжок перетворення (сприйняття докільля нашою системою відчуттів, опрацювання цієї інформації на нейронному рівні, особливості трансляції її на мовний рівень) ми не зможемо принципово наблизитися до таїн мовної організації і тим більше - будувати адекватні моделі.

В цьому плані дуже доречним буде згадати загальну оцінку класичної лінгвістики, озвучену одним із фундаторів напрямку семантичних

досліджень “Смисл - Текст” Перцовим Н.В. у лінгвістичному альманасі за 1996 рік [10]. Там він підкреслює дві головні вади класичної лінгвістики. Перша теза стосується снобізму класичної лінгвістики і зводиться до того, що коли мовленнева діяльність людини постає як одна з найскладніших форм її інтелектуальної діяльності, то лише виключно лінгвісти здатні її досліджувати і можуть вказувати всім іншим фахівцям, що і як робити в цьому напрямку; інша теза стосується того, що незважаючи на високі амбіції класичної лінгвістики її апологети – не здатні, не в змозі, не хочуть і бояться занурюватися в сфери досліджень інших помежованих наукових напрямів, і тому не враховують і не використовують їх результати, які відкривають нові горизонти пізнання мовленневої діяльності людини.

Звичайно, це – лише узагальнена характеристика сучасної парадигми класичної лінгвістики; проте є ціла когорта дослідників, які розуміють наскільки складним є шлях до появи кінцевого продукту мовленневої діяльності – мовного матеріалу і що потрібно враховувати на цьому шляху.

Найперше, це системотворчі засади Л.В. Щерби щодо дослідження мовленневої діяльності людини, які озвучені в його програмній роботі “О трояком аспекте языковых явлений и эксперименте в языкознании” [2] і зводяться до наступного: мовленнева діяльність людини актуалізується ІМС, яка може функціонувати в режимах синтезу (говоріння) та аналізу (розуміння) мовного матеріалу (ММ); процедури синтезу та аналізу ММ – це явища одного порядку складності, для розуміння та дослідження яких *обов’язковим є звернення до психонейрофізіології*; складниками ІМС постають, з одного боку наша мовна компетенція щодо мовної організації (граматика та словники), а з іншого – весь мовний матеріал, де зберігаються знання як стосовно залежностей зовнішнього світу (середовища нашого життя-буття), так і знання, власне, про самого себе (суб’єктивне “Я” або “ЕГО” кожного індивіда).

Звегінцев В.А., сприйнявши парадигму класичної лінгвістики через призму всіх європейських мов, дуже прискіпливо досліджує структурну організацію мови висловивши важливу тезу стосовно структурної організації мовного матеріалу: мовна організація визначається системою різнопланових одиниць кількох рівнів з певною ієрархією, де одиниці нижнього рівня мають засоби формування одиниць наступного рівня. Так елементи морфології формують синтаксичний рівень повідомлення; засоби формування окремого повідомлення мають відповідні елементи формування абзацу, забезпечуючи зв’язність тексту; абзаци, в свою чергу, постають структурно-організаційними компонентами тексту тощо. Інша його важлива думка стосується особливостей організації мовної діяльності людини, взагалі. Простеживши десь двадцятирічний етап розвитку всіх європейських мов (він довгий час був відповідальним редактором серії “Новое в зарубежной лингвистике”) Звегінцев В.А. зрозумів складність мовної організації зокрема і мовленневої діяльності взагалі, синтезувавши тезу: мовленнева поведінка (діяльність) людини – це настільки

складне інтегральне явище, що досліджуватись воно може лише з позицій урахування **триєдності мови, мислення та зовнішнього світу** [11]. Ця теза, практично, перегукується з тезою Л. Щерби: мова – це об’єкт дослідження класичної лінгвістики, мислення – зачіпає особливості формування ММ на психофізіологічному рівні, а зовнішній світ (точніше – знання про зовнішній світ) постають обов’язковою умовою актуалізації мовленнєвої діяльності на етапах синтезу та аналізу мовного матеріалу.

Напрочуд конструктивно на шляху дослідження мовної діяльності вбачається позиція Городецького Б.Ю., що озвучена в передмові до випуску збірника “Компьютерная лингвистика” [12] із серії “Новое в зарубежной лингвистике”. Наведемо повністю цитату, щоб точно передати зміст “Многие беды в языкознании связаны с тем, что до сих пор язык считается формой отображения мысли, а не формой представления и организации знаний”. Для нас це – ключова позиція (дороговказ) на шляху дослідження та моделювання мовленнєвої діяльності людини; тобто, **головна функція мови**, а отже і складного мовотворчого процесу (синтезу повідомлення) – **це формування знань**. Поки інформація знаходиться ще на образному рівні і не озвучена (не вербалізована), вона не уявляє собою знання. Знання з’являється лише тоді, коли окремим складникам ситуації образного рівня присвоюються спочатку певні мовні категорії, які потім пов’язуються відповідними структурно-функціональними залежностями.

А тепер кілька слів стосовно моделювання поки що окремих проявів мовленнєвої діяльності, які закладаються в сучасні інформаційні технології, що орієнтовані на опрацювання природно-мовної інформації. За визнанням Перцова М.В. [10] у справу моделювання окремих напрямів мовленнєвої діяльності більш значний внесок зробили не лінгвісти, а фахівці інших напрямів: програмісти, працівники інформаційних служб, фахівці когнітологи, кібернетики тощо. Це пояснюється тим, що всі люди як носії та користувачі мови, особливо фахівці вказаних напрямів (що гарно володіють інструментарієм моделювання) мають своє бачення вирішення вказаних проблем, яке визначається інтуїтивним характером розуміння мовної організації та необхідністю отримати хоча б якийсь програмний продукт. Більш переконливий аргумент на користь такого підходу зводиться до того, що головна теза класичної лінгвістики **“Лише через мову ми пізнаємо людину”** тепер замінюється більш продуктивною та конструктивною, на наш погляд, тезою **“Лише пізнавши добре людину (зокрема – як працює наш мозок) ми зможемо краще зрозуміти особливості мовної організації та мовленнєвої діяльності людини”**.

Одним з таких фахівців був Мельніков Г.П. – фізик за фахом (канд. фіз.-мат. наук), у подальшому, завідувач кафедри російської мови у міжнародному університеті ім. Патріса Лумумби, доктор філологічних наук (посмертно), який побудував гарну віртуальну модель мовленнєвої діяльності людини, де на високому формальному рівні досліджуються окремі етапи синтезу мовного повідомлення [3]. Важливим для нас буде його

висновок (хоча і не зовсім чітко окреслений), що коли ІМС охоплює в своїй сукупності знання про мовну організацію (ЛП) та знання про довкілля, що зберігаються у (БЗ), то і розбудова моделі мовленнєвої діяльності повинна йти шляхом формування моделей M_1 (як моделі ЛП) і M_2 (як моделі нашого оточення, що зберігається в БЗ). Лише після формування моделей M_1 та M_2 і замикання їх одна на одну через відповідний інтерфейс, тобто, лише після створення моделі інформаційного ядра, ми зможемо вийти на етап моделювання окремих проявів мовленнєвої діяльності шляхом адаптації ядра ІМС вже під конкретні інформаційні технології.

Висновки

Отже стратегічний напрям дослідження мовленнєвої діяльності вбачається наступним.

1. На шляху дослідження МД інтегруються сучасні здобутки всіх помежованих напрямів, які тим чи іншим чином стосуються мовленнєвої діяльності людини; це будуть нейрофізіологія, психологія, філософія, інформатика, когнітологія тощо.
2. Чітко аналізуються здобутки та недоліки класичної лінгвістики і з урахуванням здобутків у помежованих напрямках формується нове бачення структурної організації мови, що постає обов'язковою умовою створення програмного продукту під ЛП.
3. З позицій системної організації мови формується ідеологія розбудови природно-мовної бази знань (БЗ), наповнення якої буде модель середовища.
4. Після формування моделей ЛП та БЗ з відповідним інтерфейсом можемо вже говорити про створення у повному обсязі моделі ІМС.
5. Наступний етап пов'язаний вже буде з адаптацією моделі ІМС під конкретні інформаційні технології опрацювання природно-мовної інформації.

Література

1. Кисленко Ю. От мысли к знанию (нейрофизиологические основания) – Монография - Киев : Издательство “Український літопис”, 2008. – 102 С.
2. Щерба Л.В. О трояком аспекте языковых явлений и эксперименте в языкознании. // Языковая система и речевая деятельность. – М., 1974.
3. Мельников Г.П. Системология и языковые аспекты кибернетики М., “Сов. Радио”, 1978. – 368 с.
4. Гаврилова Т.А. , Хорошевский В.Ф. Базы знаний интеллектуальных систем: Уч. пособ. для студ. вузов – СПб.: ПИТЕР, 2001. 384

5. Звегинцев В.А. История языкознания XIX – XX веков в очерках и извлечениях – ч. II. - М.: Просвещение, 1965.
6. Пиотровский Р.Г. - Лингвистические уроки машинного перевода, Вопросы языкознания, 4, 1985. – с. 18-27.
7. Астахова Л.И. Предложение и его членение /прагматика, семантика, синтаксис/ Днепропетровский ГУ, 1992.
8. Грамматика современного русского литературного языка. Наука, М., 1970, (Грамматика-70).
9. Золотова Г.А. Коммуникативные аспекты русского синтаксиса. – М.: Наука, 1982.
10. Перцов Н.В. О некоторых проблемах современной семантики и компьютерной лингвистики // Московский лингвистический альманах, 1996, - Вып. 1, с. 9-66.
11. Звегинцев В.А. Предложение и его отношение к языку и речи Изд. МГУ, 1976. – 308 с.
12. Городецкий Б.Ю. Компьютерная лингвистика: моделирование языкового общения. (Пер. с англ. /сост., ред. и вступ. ст. Б.Ю. Городецкого). - Серия "Новое в зарубежной лингвистике", Вып 24. – М.: Прогресс, 1989.- 432 с.

Отримано 06.03.2010 р.