

Ілля Данилюк
(м. Вінниця)

МОДЕЛЬ МОВНОЇ ОСОБИСТОСТІ НА РІВНІ ПУНКТУАЦІЇ¹

Вчені пропонують кілька визначень поняття *мовної особистості* [Воркачев, 2001]. З одного боку, мовною особистістю називають людину як носія мови, взяту в аспекті її здатності до мовленнєвої діяльності, тобто це по суті комплекс психофізичних властивостей індивіда, що дозволяє йому генерувати і сприймати мовлення, – *особистість мовленнєва*.

Мовною особистістю називають також сукупність особливостей вербальної поведінки людини, що використовує мову як засіб спілкування, – *особистість комунікативна*.

Крім того, мовною особистістю називають закріплений переважно в лексичній системі базовий національно-культурний прототип носія певної мови, щось на кшталт «семантичного фоторобота», який складається на основі світоглядних інтенцій, ціннісних пріоритетів і поведінкових реакцій, зафіксованих у словнику – *особистість словникова*, або *етносемантична*.

А.П. Загнітко тлумачить мовну особистість як сукупність здібностей і характеристик людини, що зумовлюють продукування і розуміння особистістю текстів, що різняться між собою: 1) ступенем мовно-структурної складності; 2) глибиною, послідовністю і точністю вияву дійсності; 3) комунікативною настановою; 4) естетичним навантаженням; 5) психологічним ефектом тощо [Загнітко, 2016]. Тому мовна особистість – це поєднання в особі мовця: а) мовної компетенції, б) прагнення до творчого самовияву, в) вільного, автоматичного здійснення усебічної мовленнєвої діяльності.

В. В. Красних [Красных, 1998, с. 8] пропонує розмежовувати поняття, пов'язані з особистістю: *мовець* – особистість, одним із видів діяльності якої є мовленнєва діяльність (що охоплює як говоріння, так і процес сприймання мовлення); *мовна особистість* – особистість, що проявляє себе в мовленнєвій діяльності, володіє певною сукупністю знань та уявлень; *мовленнєва особистість* – особистість, яка реалізує себе в комунікації, вибирає та реалізує ту чи іншу стратегію і тактику спілкування, вибирає та використовує той чи інший репертуар засобів (як власне лінгвістичних, так і екстралінгвістичних); *комунікативна особистість* – конкретний учасник конкретного комунікативного акту, який реально діє в реальній комунікації.

¹ Дослідження виконано в межах фундаментального наукового проекту «Об'єктивна і суб'єктивна мовносоціумна граMATика: комунікативно-когнітивний та прагматико-лінгвокомп'ютерний виміри» (0118U003137) Донецького національного університету імені Василя Стуса Міністерства освіти і науки України

Ю. М. Караулов описує структуру мовної особистості як сукупність трьох рівнів: а) *семантичного* – цей рівень передбачає для носія нормальне володіння повсякденною мовою, для дослідника семантичний рівень є традиційним описом формальних засобів вираження певних значень; б) *лінгвокогнітивного* – цей рівень передбачає виявлення, встановлення ієрархії смислів і цінностей у картині світу мовної особистості, в її тезаурусі, охоплює інтелектуальну сферу особистості, даючи досліднику вихід через мову, через процеси говоріння і розуміння – до знання, свідомості, процесів пізнання людини; в) *прагматичного* – рівень, який включає цілі, мотиви, інтереси та інтенції. Цей рівень забезпечує в аналізі мовної особистості закономірний і обґрунтований перехід від оцінок її мовленнєвої діяльності до осмислення мовленнєвої діяльності у світі [Караулов, 1987].

Отже, поняття мовної особистості має різні рівні абстракції, що передбачає кілька аспектів її вивчення. Початковий ступінь абстракції – конкретна мовна особистість, наприклад, письменника, вченого, політика тощо (пор. [Мацько, 2013]). Так само мовною особистістю можуть називати особистість пересічного, звичайного носія мови, особливості якого проявляються в продуктах його мовної діяльності.

Наступний рівень абстракції – колективна мовна особистість, наприклад, мовна особистість в гендерному аспекті або мовна особистість певної професійної групи ([Іваницька, 2012]).

Найвищий ступінь абстракції – мовна особистість народу, національна мовна особистість, наприклад, українська мовна особистість ([Редько, 2014]).

Загалом, дотримуючись тенденцій сучасної лінгвоперсонології, ми розглядаємо мовну особистість як конкретного індивіда.

Нарешті, у дослідженні мовної особистості виділяють два підходи: 1) дедуктивний (від особистості) і 2) індуктивний (від мови).

За першого підходу вивчають особистість, її комплексний портрет, включно з мовленнєвими характеристиками, які доповнюють, підтверджують те, що було виявлено психологами, соціологами, фізіологами тощо. Для такого підходу характерним є, що позамовні типи подано апіорі і лінгвіст шукає їхні мовні відмінності.

У межах другого підходу описують мову, текст, мовленнєву поведінку в персонологічному аспекті. Тут як предмет вивчення виступає мова, що характеризується в аспекті особистісних властивостей. За такого підходу типи мовної особистості виділяються за результатами дослідження продуктів мовленнєвої діяльності носіїв мови. Так, наприклад, лінгвоперсонологічного характеру набувають синтаксичні одиниці, частини мови, лінгвокогнітивні стилі репродукування наукового тексту тощо.

Для моделювання процесу діяльності мовної особистості можна запропонувати систему координат трьох категорій, запропонованих у [Геодакян, 1970]: *простір здатностей*, *простір можливостей* і *простір*

реалізацій. Перший з них – це сукупність внутрішніх ступенів свободи системи, які утворюють певний багатовимірний простір. Це перелік усього, що система здатна робити, перелік внутрішніх потенційних програм системи. Простір можливостей – це сукупність зовнішніх ступенів свободи, які так само утворюють багатовимірний простір, це перелік усього того, що системі дають змогу реалізувати умови середовища, у якому вона функціонує. Точки, які належать одночасно до простору здатностей і простору можливостей, утворюють третій простір – простір реалізацій, який характеризує поведінку системи.

Якщо спроектувати таку модель координат на мовну особистість як систему, то простір здатностей конкретної мовної особистості утворений набором мовних одиниць (фонем, морфем, слів тощо) та правил їх поєднання, які зберігаються в мовній свідомості (пам'яті) людини. Простір можливостей для мовної особистості – це сукупність специфічних дискурсивних чинників, які формують межі вільного використання людиною мови для досягнення комунікативних цілей. Нарешті, простором реалізацій мовної особистості є множина породжених нею текстів, у яких відбивається взаємодія, перетин простору здатностей і простору можливостей.

З трьох наведених просторів останній може бути об'єктом емпіричного дослідження з використанням методів машинного навчання. Іншими словами, маючи в розпорядженні корпус текстів, породжених мовною особистістю, можна встановити конкретні елементи її простору здатностей, а інформація про біофізіологічні, психологічні, соціально-статусні характеристики мовної особистості, її біографічні дані дають змогу реконструювати й простір можливостей.

Простори здатностей і реалізацій мовної особистості можуть бути сегментовані за різними критеріями, зокрема і з традиційним для мовознавства поділом на лексичний та граматичний рівні, на рівень орфографії (тобто особливостей дотримання мовною особистістю певних орфографічних норм) та орфоепії (дотримання норм вимови) тощо. Одним з таких рівнів є пунктуаційний, тобто вживання мовною особистістю розділових знаків. З одного боку, їх вживання регулюється правописом, з іншого – мовна особистість завжди обирає, як і де ставити розділові знаки, чи буде мовлення так, аби цей вибір був для неї максимально простим. На використання розділових знаків може впливати як знання чи незнання чинних норм, так і внутрішнє бажання мовної особистості. Наприклад, поет Олександр Косенко в окремих творах зі збірки «На сьомий день» взагалі уникає пунктуації.

У межах вивчення мовної особистості Юрія Шевельова (Шереха) було зібрано ємний корпус текстів. Для моделювання мовної особистості вченого методом суцільної вибірки було сформовано підкорпус з речень з комами, з мінімальною довжиною 3 слова. Характеристики підкорпусу: 4422266 знаків,

656033 токенів, 36958 речень, у яких вжито 88492 коми. Було висунуто гіпотезу про можливість створення прогностичної моделі пунктуаційного рівня мовної особистості Юрія Шевельова, здатної відтворювати пунктуацію (вживання ком) у тестових реченнях (текстах), де її спеціально видалено.

Було використано 3 методи з категорії машинного навчання: 1) метод опорних векторів, у якому модель прогнозує, чи треба поставити кому між 2 словами, спираючись на а) 2 останні літери попереднього слова і наступне слово; б) ціле попереднє і наступне слово; 2) двонаправлена рекурентна штучна нейронна мережа (RNN): модель спирається на проходження тестового речення зліва направо і справа наліво, прогножуючи, чи треба і на якій саме позиції вставити кому; 3) згортоква штучна нейронна мережа (CNN): модель спирається на проходження тестового тексту рамкою у 15 слів, прогножуючи, чи потрібно ставити кому між 7-м і 8-м словом. Усі три методи показали значно вищу за випадкову прогностичну силу.

Зокрема, в моделі, побудованій з використанням RNN, побудованій, як описано в [Tilk, Alumäe, 2016], усі розділові знаки в тренувальній вибірці були замінені на штучні токени: ?QUESTIONMARK, !EXCLAMATIONMARK, ,COMMA, -DASH, :COLON, ;SEMICOLON, .PERIOD. Відповідно, декодер генерує ланцюжки зі слів вхідного речення, додаючи відповідні токени, порівняйте результат:

кінець кінцем видавці справді стали домагатися ,COMMA щоб дюма писав сам .PERIOD

тоді йому пощастило знайти людину з таким же письмом ,COMMA як у нього ,COMMA і ця людина переписувала романи ,COMMA а дюма приносив їх ,COMMA хвалячися ,COMMA що це вже він нарешті сам написав .PERIOD

безперечно ,COMMA дюма був у романах такий же спритний комбінатор ,COMMA як і в комерції .PERIOD

але назвати цю фабрику безідейного й безстильового читва « величчю світової літератури » ?QUESTIONMARK

достоевський писав про дюма :COLON « саме тому дюма і не мистець ,COMMA що він не може стриматися в своїй розгнущданій фантазії від перебільшених ефектів .PERIOD

але навітьо смарагдовий флякон для отрути ?QUESTIONMARK

навіщо та споживна речовина ,COMMA що її однією щипкою він міг наїстися на кілька днів ?QUESTIONMARK

звичайно ,COMMA є фізична ймовірність відшукати в природі великий смарагд ,COMMA придатний на флякон ,COMMA але ж треба знати й міру ,COMMA треба вміти стриматися своєчасно .PERIOD

Отже, використання методів машинного навчання дає змогу змоделювати різні рівні мовної особистості, спираючись на простір реалізацій, і застосувати її для реконструкції ймовірного простору реалізацій, а також відтворення простору здатностей конкретної мовної особистості.

Література

1. Tilk O., Alumäe T. Bidirectional recurrent neural network with attention mechanism for punctuation restoration // Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH. 2016. Т. 08-12-Sept. № September. С. 3047–3051.
2. Воркачев С.Г. Лингвокультурология, языковая личность, концепт: становление антропоцентрической парадигмы в языкознании // Филологические науки. 2001. Т. 1. № 1. С. 64–72.
3. Геодакян В.О. Организация систем - живых и неживых // Системные исследования 1970. Ежегодник. Москва: Наука, 1970. С. 49–62.
4. Загнітко А.П. Лінгвоіндивідуації та лінгвоіндивідуалізації в епістолярному дискурсі: закономірності реалізації мовної особистості // Лінгвокомп'ютерні дослідження. 2016. Т. 9. С. 121–143.
5. Іваницька М.Л. Мовна особистість перекладача як чинник впливу на продукт перекладу (на матеріалі роману Б. Леберта Крейзі) // Проблеми семантики, прагматики та когнітивної лінгвістики. 2012. № 21. С. 147–159.
6. Караулов Ю.Н. Русский язык и языковая личность. : УРСС, 1987. 262 с.
7. Красных В.В. Виртуальная реальность или реальная виртуальность?: человек, сознание, коммуникация. Москва: Dialog-Mgu, 1998. 352 с.
8. Мацько Л.І. Мовна особистість Андрія Малишка як лінгвокультурологічний феномен національного виховання у світлі лінгвофілософії // Науковий часопис НПУ імені М. П. Драгоманова. Серія 16 Творча особистість учителя проблеми теорії і практики. 2013. № 20. С. 141–146.
9. Редько Є.О. Українська мовна особистість і арго (на прикладі семантичних назв осіб за спорідненістю і свояцтвом) // Науковий часопис НПУ імені М. П. Драгоманова. Серія Філологічні науки (мовознавство і літературознавство). 2014. № 5. С. 212–218.

Наталія Дарчук
(*м.Київ*)

ДЕРЕВО ЗАЛЕЖНОСТЕЙ ЯК ПАРАМЕТР АВТОРСЬКОГО СТИЛЮ

У багатьох системах автоматичного опрацювання тексту і вітчизняних, і зарубіжних використовується спосіб формального синтаксичного аналізу, який базується на граматиці залежностей (ГЗ), а структура речення має вигляд дерева залежностей (ДЗ). В основі ГЗ лежить синтаксис Теньєра [3]. Будова речення уподібнюється молекулі, складається з атомів – слів, пов'язаних між собою. Принципи такі: