

### Системи анотування просодії китайської мови

*Гобова Є.*

Із пошкваленням розвитку цифрових технологій акустичні дослідження просодії вийшли на новий рівень. Завдяки новітнім досягненням у цій сфері для широкого загалу стали доступні різноманітні засоби, що дають змогу проводити аналіз просодичного матеріалу усного мовлення з доволі високою точністю. Проте, окрім питань пов'язаних з суто технологічною стороною цього процесу, актуальною стала проблема уніфікованого підходу до анотування мов із різними просодичними атрибутами. Так навіть близькоспоріднені мови вимагають кардинально відмінних підходів до опису та транскрибування.

Першим етапом розвитку та впровадження системи анотації та транскрибування стало створення баз даних призначених для неавтоматизованої обробки спеціалістами переважно лінгвістичної галузі. Саме на цьому етапі були закладені основні риси системи ToVI – чотири мінімальні рівні анотування, основні графічні символи-позначки. Окрім слів, усне мовлення оперує такими інформативними надсегментними елементами як тривалість, інтонація, наголошеність, поділ на мовні одиниці, якість звуку, ритм тощо. У сукупності вони складають просодію мовлення.

Поміж інших спроб адаптації системи просодичної анотації ToVI до окремих мов світу, створено кілька системи для китайської мови. Зокрема першою системою стала C-ToVI, вперше випущена 1996 року [Beckman 1997, 10]. Через фактичну відсутність робочої інтонаційної теорії мовлення китайської мови, розробники системи були змушені використовувати результати різних просодичних досліджень. Створення цієї системи засновувалося на таких принципах: по-перше, головним призначенням цієї системи є розмітка тональних варіацій, інтонації, наголосу та просодичних структур, що мають лінгвістичні функції. Так, наприклад, тональна коартикуляція між складами не анотується, натомість анотується тональна коартикуляція, що виникла під дією наголосу. По-друге, анотуються якісні риси просодії, на відміну від кількісних, таких як тривалість та інтенсивність. Крім того, “непевні” місця анотації відмічаються задля уникнення неправильного розуміння іншими користувачами системи. Оскільки транскрибування проводиться не одним користувачем, між ними має існувати узгодженість та послідовність дій.

Найбільш очевидною відмінністю C-ToVI від оригінальної системи анотації, створеної на основі специфіки англійської мови, можна назвати відносно “багатошаровість” її структури. В той час як для E-ToVI базова кількість рівнів анотації становить чотири, включаючи “тони”, “паузи”, “слова” та категорію “різне” для особливих явищ, часто екстралінгвістичного характеру [Beckman 1997, 38], то для C-ToVI її автори пропонують аж вісім рівнів. Серед них: “ієрогліфи”, “піньїнь”, фонетична сегментація “ініціали-фінали”, категорія “різне”, “тип речення”, “паузи”, “наголоси” та “експресивність”. Звичайно ж не завжди є потреба у використанні всіх цих рівнів, тому найчастіше у вжитку цієї системи найчастіше використовуються практично ті самі ж рівні, що і в оригінальній системі ToVI, виключення становлять ті рівні, що безпосередньо пов'язані з графічною специфікою китайської мови.

Також існує система анотації під назвою M-ToVI, що ще іноді називається “всекитайською” (Pan-Mandarin) через намір її авторів створити систему анотування просодичної інформації не тільки у стандартизованій китайській мові путунхуа але й у її місцевих різновидах та, в перспективі, для всіх діалектів китайської мови. Щоправда, думки самих розробників розбігаються щодо можливості та доцільності її застосування для таких значних діалектних груп як У, Юе та Мінь [Lee 2007, 12]. Варто відмітити, що для основних груп діалектів також створюються та вже створені відповідні системи [Chen 2000, 58].

Рівень “пінїнь” містить у собі фонетичні позначення за допомогою стандартного китайського алфавіту пінїнь, в яких тони позначаються не діакритичними знаками, а цифрами від 0 до 4, де 0 – нейтральний тон, а 1-4 відповідно тони від першого до четвертого. Таке маркування тонів обумовлене технічною специфікою – програми, що використовуються для анотування найчастіше не підтримують діакритичні позначки.

Рівень “ієрогліфи” містить у собі запис висловлювання оригінальною ієрогліфікою. Проте, на відміну від інших рівнів, в цьому символі використовуються не як теги, чи позначки, а як додаткова інформація, що супроводжує процес анотування та аналізу. Тобто ієрогліфічний запис слугує семантичною приміткою не прив’язаною напряму до інших рівнів, чи візуального вираження анотованого уривку мовлення. Рівень “ініціалі-фінали” за допомогою символів МФА чи пінїньну репрезентує компоненти складів.

На рівні “тонів та інтонації” за допомогою системи позначок маркуються зміни регістру. Через неусталеність інтонаційної граматики китайської мови інтонаційна транскрипція застосування деяких позначок може бути неоднозначним на різних етапах обробки та аналізу мовленнєвої інформації. Саме на цьому рівні реалізується одне з основних призначень системи анотування. Через це між різними мовно-спеціалізованими системами ToVI головні відмінності у анотаційних засобах відносяться до рівня тонів та інтонації. Наприклад, у системі C-ToVI впроваджено наступні символи:

*H-L, L-H, H-H, L-L*: тональні властивості чотирьох тонів стандартної китайської мови;

*H, L*: тональні властивості нейтрального тону стандартної китайської мови;

*L%, H%, %L, %H*: кінцевий граничний тон, початковий граничний тон;

*&*: перехідний тон;

*^, !, ^^, !!*: підйом, широкий підйом, опущення, широке опущення;

*Re^(), Re^^(), Re!(), Re!!()*: зміна висоти регістру, де дужками позначаються межі зміни;

*Ra^(), Ra^^(), Ra!(), Ra!!()*: глобальна варіація висоти регістру, де дужками позначаються межі варіації;

*H^^, H^, L!* : локальна варіація висоти регістру.

Іншим специфічним рівнем анотування є рівень «паузи». Залежно від загального типового темпу мовлення, притаманного кожній конкретній мові, ієрархія позначень може суттєво відрізнитися. Паузи позначаються між двома складами тоді, коли транскрибувальник здатний їх сприйняти на слух. Місце та тривалість пауз може змінюватися у різних мовців, що начитують один і той самий текст, та навіть у одного мовця, що читає той самий текст через невеликі періоди часу. Автори системи C-ToVI пропонують використовувати для анотації китайської мовлення паузи п’яти категорій відповідно до їх фактичної тривалості:

- Індекс паузи = 0: мінімальна пауза між складами, найчастіше це пауза у просодичному слові;

- Індекс паузи = 1: просодичні межі слова;
- Індекс паузи = 2: межі малої просодичної фрази;
- Індекс паузи = 3: межі великої просодичної фрази (інтонаційної фрази);
- Індекс паузи = 4: межі просодичної групи;
- 1р, 2р та 3р позначають аномальні паузи, наприклад спричинені ваганням, кашлем тощо;
- “?”: невизначеність паузи.

Між іншим, у системі M-ToBI наряду з індексацією пауз обумовлені типові значення індексів рівнів “наголос” та “тон”, що є закономірними для них:

- Елемент з індексом паузи (далі – VI від *Break Index*) = 0 зліва або справа вимагає наголосу зі значенням 0 або 1;
- Після елемента з VI=2 має бути наголос зі значенням щонайменше 2;
- Межі просодичної групи, з VI=4 мають позначатися відповідними символами для граничних тонів.

Разом з тим, у цих двох системах дещо відрізняються класифікації пауз. Наприклад, у M-ToBI вона нараховує на одну ступінь більше – тут ще надається окрема позначка паузи для “переведення дихання” VI=5 (*breath group*). Проте така деталізованість виявилася надмірною вже на перших етапах практичного випробовування системи – як з’ясувалося, для транскрибувальників було проблематичним розрізнення VI=4 та VI=5 [Lee 2007, 261]. До того ж, така докладність абсолютно не принципова у більшості випадків користування системою.

Специфічним рівнем, характерним саме для китайської мови, можна назвати рівень “сандхі”, запропонований до використання розробниками Pan-Mandarin-ToBI [Lee 2007, 246]. За прикладами з діалектних варіантів китайської мови були відстежені відмінності у функціонуванні змішання лексичних тонів. Таким чином, доцільним виявилось також введення окремого рівня для нього. Для запису значення рівня “сандхі” прийнято використовувати так звану “цифрову систему Чао”, тобто послідовність двох чи більше цифр ряду від 1 до 5, що схематично позначає загальний вигляд реалізованого лексичного тону.

Крім того, введені додаткові суто функціональні позначки. Всі акустичні події, що не належать до лінгвістичних явищ та істотно не впливають на процес опрацювання матеріалу, прийнято фіксувати на рівні під назвою “різне” за допомогою відповідних умовних слів-позначок, наприклад: “кашель”, “сміх”, “шум” тощо.

При графічному позначенні в одній зі спеціально створених програм (Praat, waves тощо) можна створювати рівні кількох типів: такі, що вказують на значення конкретної обраної точки за певними параметрами та такі, що позначають межі одиниці категорії, обраної для конкретного рівня. Так, наприклад рівень “висоти звучання” та “наголосу” зазвичай при створенні встановлюють “точковим” (*point tier*), а такі рівні як “піньїнь” та решта подібних встановлюють “інтервальними” (*interval tier*).

Серед найбільш типових відмінностей між M-ToBI та C-ToBI слід відзначити саме відсутність рівня “сандхі” в останньому [Lee 2007, 126]. Крім того, C-ToBI пропонує введення окремого рівня для вказівки на приналежність мовлення до діалектного за допомогою відповідних позначок. Достеменно важко сказати, наскільки це доречно, враховуючи, що ця система створювалася саме для путунхуа, проте в цьому випадку можливі різні тлумачення самого поняття “діалект”, оскільки на рівні з ним у даній статті в аналогічному контексті функціонує термін “акцент”, що в принципі не є тотожним поняттям. Цілком можливо, що цей рівень було запропоновано для позначення виключних, а не закономірних випадків, тобто не для систематичного опису діалектів китайської мови, на відміну від M-ToBI.

---

Для інших різновидів анотації мовного матеріалу, наприклад синтаксичної, частини мови транскрибуються аналогічно до систем синтезу мовлення. Серед інших нетипових рівнів анотування – емоції та ставлення, що часто зводяться в одному рівні “ефективність”. Емоції означають внутрішні почуття мовця – радість, злість, здивування, агресивність тощо. Ставленням описуються зовнішні прояви поведінки мовця – дружність, співчуття, вибачення і т. п. На даний момент створено програму Feeltrace, що слугує допоміжним засобом для впровадження маркування афективної інформації мовних корпусів у різних часових проміжках відповідно до сприйняття слухачем. Проте, наразі не створено жодної послідовної системи такого анотування для окремої мови.

Укладачі базового варіанту ToBI розрахованого на англійську мову, перш за все наголошують на тому, що цей інструмент створювався головним чином для опису ритму, або чергування різною мірою наголошених складів, принципів групування просодичних слів у більші одиниці та інтонаційних моделей, що складаються з сукупності фразових акцентів, різниці у висоті звучання та граничних тонів, функціонування, що уможлиблюють проведення опису сандхі тонів, зокрема особливо складних випадків послідовностей третього тону.

Натомість деякі типові аспекти просодії не тільки не мають безпосереднього усталеного способу запису у цій системі, але й разом з тим не передбачені укладачами. Серед них – темп та лексичний наголос. Причин для цього може бути кілька: по-перше, ці категорії мають лише дотичне відношення до першочергового призначення системи, по-друге, у випадку з темпом занадто вузькі можливості для маркування цієї категорії, оскільки темп та тривалість звучання некоректно записувати позначками лише з кількома варіаціями значення, що є достатньо типовим та зручним для інших категорій. По-третє, у випадку з лексичним наголосом, немає потреби у його фіксації, оскільки словниковий варіант наголосу має бути відомий слухачеві апріорі, а відхилення від норми все одно фіксуються на інших рівнях.

Разом з тим існують й інші погляди на доцільність використання системи анотації ToBI в межах групи мов чи діалектів однієї мови, оскільки окрім фіксацій тональних малюнків, ToBI також фіксує моменти переходу між двома суміжними словами, тобто транскрибування фактично відбувається на фонологічному а не фонетичному рівні [Yip 2002, 180]. Таким чином, певний набір інструментів-рівнів, що використовується для однієї мови чи діалекту може бути непридатним для іншого, навіть якщо відмінності між ними на перший погляд незначні. Тому випадки типологічного чи порівняльного опису фонології кількох мов доцільно проводити із використанням інших транскрибувальних систем [Cummins 1999, 25]. До послуг дослідників наразі запропоновано, наприклад, таку універсальну систему як INSINT – International Transcription System for Intonation [Hirst, Di Cristo 1998, 14]. З одного боку, вона має доволі вузьке призначення – виключно в рамках інтонації, з іншого, до її переваг можна віднести можливість одночасного транскрибування великої кількості мов за допомогою доволі простого та інтуїтивно зрозумілого способу анотування. Зокрема цю систему широко вживають при описах інтонаційних систем мало досліджених чи рідкісних мов. Також, порівнюючи цю систему з ToBI, очевидно, що INSINT фіксує виключно інформацію по висоті звучання, при тому, що ToBI окрім висоти транскрибує й інші просодичні дані, зокрема й міру сполученості окремих просодичних елементів. При цьому така з першого погляду «невизначена» структура ToBI, насправді є перевагою в тому плані, що немає потреби використовувати одну громіздку систему фонетичного запису для всіх мов та діалектів, оскільки цілком достатньо, щоби

---

кожен різновид системи мав свою послідовну систему анотування. Особливо зручним це виявляється при введенні рівня “фонемі” чи для мало описаних мов.

Специфіка запису INSINT головним чином полягає у використанні міток ↑, ↑↑, ↓, ↓↓ для позначення зростаючої інтонації, спадаючої інтонації та посиленних варіантів обох інтонацій відповідно. Очевидність та простота цього способу анотування є значною перевагою в контексті одночасного дослідження цілої множини мов та діалектів. Проте в цьому полягає і деяка примітивність цієї системи – опис при цьому занадто поверхневий та прямолінійний. Тому цю систему прийнято називати «фонетичною» на протигагу “фонологічній” ToVI через відсутність гнучкості та можливостей деталізації.

Перш за все, варто відзначити, що самі автори стверджують про недоцільність використання одного різновиду ToVI для кількох мов, чи навіть кількох діалектів однієї мови [Beckman 2006, 31]. Разом з тим, вони визнають необхідність використання принаймні чотирьох основних рівнів разом із записом F0 та аудіо-файлом у будь-якому специфічному різновиді цієї системи. Більше того, заохочується створення нових рівнів, відповідно до побажань та потреб дослідників. Таким чином вже було створено чимало можливостей для транскрибування дискурсу, синтаксичної сегментації тощо. Фактично на основі ToVI було створено систему транскрибування жестів та міміки – Anvil, що працює з файлами візуального відтворення.

З іншого боку існують спроби відмовитися від обов’язкового мінімального набору інструментів. Зокрема Cantonese-ToVI, створений для транскрибування кантонського діалекту, не включає рівня «Наголос», натомість введено рівень “Романізація”, або “Транслітерація”. Така заміна цілком прийнятна з огляду на специфіку цієї мови: наголос тут не є принциповою складовою мовлення і практично не виділяється з загального мовного потоку, змішуючись зі складною системою тонального малюнку. Разом з тим рівень транслітерування за допомогою символів латинського алфавіту цілком доречний виходячи з суто технічних міркувань – деяке програмне забезпечення, за допомогою якого проводяться дослідження, може не підтримувати як дореформене, так і спрощене ієрогліфічне письмо, в той час як романізація є доступним та доволі зручним засобом запису.

Відкритим лишається питання доцільності використання рівня «Наголос» для різних китайських мов. Як вже було сказано вище, для кантонського різновиду системи ToVI та кантонської мови як такої не прийнято виділяти таке поняття як наголос взагалі [Lee 2007, 432], натомість для СКМ запропоновано окремо описувати наголос, а лексичні тони відмічати разом із транслітерацією. Звичайно ж, доцільність використання тих чи інших засобів створених розробниками цих систем просодичної анотації цілком визначається користувачем відповідно до безпосередніх потреб. До того ж, окрім запропонованих рівнів та їх індексів, цілком можливо створити свої власні функціональні інструменти, чи розширити набір доступних індексів. Проте доцільним та ефективним таке доповнення буде за умови дотримання послідовності введення позначень [Dong, Lua 2000, 8] .

Також, з наведеної у “Просодичній типології” таблиці не важко зробити висновок, що серед усіх розроблених на даний момент різновидів цієї системи, M-ToVI (в даному випадку це Pan-Mandarin) має найбільш поширений та складний інструментарій, порівняно велику кількість типів індексів паузи [Lee 2007, 434-435]. Натомість кількість індексів тонів значно поступається корейській мові, зокрема у поширеності та складності сполучень. При цьому багато подібностей можна знайти між цією системою та, наприклад, ToVI неаполітанської італійської мови.

На даний момент створено кілька комп'ютерних програм для обробки, аналізу та відтворення мовлення. Серед них є й такі, що підтримують просодичну розмітку аудіо-, відео- чи текстового файлу. Найбільш доступною для широкого загалу користувачів наразі є програма Praat, яку створили Пауль Боерсма та Давід Веєнік в Інституті фонетичних наук при Амстердамському університеті. Слід зазначити, що просодичне транскрибування є лише однією з функцій цієї програми поряд із аналізом та синтезом окремих звуків. Хоча така багатосторонність програми значно ускладнює користування нею – більшість функцій потребує від користувача деяких спеціальних знань та вмінь: від базових знань фонології до вміння компілювати так звані “сценарії”-скрипти для адаптації програми під конкретні вимоги та загалом оптимальнішої роботи. Це ж стосується автоматизації процесів програми – це можливо лише частково та за умови наявності доволі високого рівня навичок користувача у сфері програмування.

Серед інших програм цього призначення варто згадати Waves від розробників Entropic Research Laboratories. Ця програма, на відміну від інших, є платною та разом з тим більш зручною у використанні навіть для менш досвідчених спеціалістів. Разом з тим, останніми роками популярність цієї програми суттєво зменшується із відсутністю можливості інтегрувати файли з інших програм анотування та відмінності у форматах опису порівняно із найбільш поширеними програмами.

Із суміжних за призначенням програм можна назвати Anvil, що також безкоштовно розповсюджується серед усіх бажаючих. Специфікою цієї програми є її спрямованість на анотування відеозаписів. В першу чергу вона була створена саме для опису особливостей жестикуляції та міміки у прив'язці до інтонаційних проявів мовлення. Можливо через те, що програма більш сучасна, в ній є дещо більше можливостей для автоматизації процесів обробки матеріалу. Разом з тим, автоматизація тут має доволі формальний характер на рівні встановлення відповідного таймінгу для субтитрів. Що характерно – ця програма сумісна із форматами файлів Praat, що тим більш важливо для дослідників, що користуються обома програмами.

## ЛІТЕРАТУРА

**Advances in Chinese Spoken Language Processing.** Edited by *Chin-Hui Lee*. – World Scientific, Singapore, 2007.

*Beckman Mary E., Hirschberg Julia, Shattuck-Hufnagel Stefanie.* The Original ToBI System and the Evolution of the ToBI Framework / **Prosodic Typology: The Phonology of Intonation and Phrasing.** – University Press, Oxford, 2006. – P. 9-55.

*Beckman Mary E.* **Guidelines for ToBI Labelling (version 3.0).** – State University Research Foundation, Ohio, 1997.

*Chen, Matthew Y.* **Tone Sandhi: Patterns across Chinese Dialects.** –University Press, Cambridge, 2000.

*Chen Weijun.* Training Prosodic Phrasing Rules for Chinese TTS Systems // **Eurospeech.** – 2001. – P. 1159-1162.

*Cummins, Fred.* **Comparing Prosody Across Many Languages.** – IDSIA, 1999.

*Dong Ming-hui, Lua Kim-Teng.* **Automatic Prosodic Break Labeling for Mandarin Chinese Speech Data.** – School of Computing at National Singapore University, Singapore, 2000.

*Hirst, Daniel and Di Cristo, Albert.* **Intonation Systems: A Survey of Twenty Languages.** – University Press, Cambridge, 1998.

*Jun Sun-Ah.* Prosodic Typology / **Prosodic Typology: The Phonology of Intonation and Phrasing.** – University Press, Oxford, 2006. – P. 430-459.

- 
- Ladd, D. Robert. **Intonational Phonology**. – University Press, Cambridge, 1996.*
- Li Aijun. Chinese Prosody and Prosodic Labeling of Spontaneous Speech / **Report of Phonetic Research**, 2000.*
- San Duanmu. **The Phonology of Standard Chinese**. – Oxford University Press, Oxford, 2007.*
- Tones and Tunes. Vol. 1-2.** Edited by *Carl Gussenhoven* – Mouton de Gruyter, Berlin, 2007.*
- Tseng Chiu-yu, Pin Shao-huang et al. **Fluent Speech Prosody: Framework and Modeling**. – Academia Sinica, Taipei, 2005.*
- Yip Moira. **Tone**. – University Press, Cambridge, 2002.*