

УЖ₂ – 20% звернули увагу на вимову слова [до Ганни], 15% на [чи Таня], 10% на [гана], 10% на [у судді], 10% на [вважала], 10% на [піддашій]; 5% на [від дому];

УЖ₃ – 25% звернули увагу на вимову слова [піддашій], 20% на [слухання], 20% на [співчуття], 20% [догани]; 10% на [гана], 10% на [у судді];

УЧ₁ – 25% звернув увагу на вимову слова [лють], 25% на [до Ганни], 25% на [читання], 20% на [у судді], 10% на [лють], 10% [безмежжі];

РЧ₂ – 50% звернули увагу на вимову слова [лють, тому, що вимовлялось *лють*], 10% на [у судді], 10% на [піддашій], 10% на [почутті], 10% на [безмежжі], 10% на [Ганна];

З експерименту видно, що форми слів *се, гаддя, безмежжі* у всіх дикторів розпізнавалися. Також була помічена вимова таких слів, як *почутті, почутті, цвіллю, чи Таня, вважала, піддашій, читання, у судді, у суді, до Ганни, догани, гана, Ганна, лють, лють, слухання, співчуття*. Найменше привернуло уваги слово *від дому*. Зовсім уваги не привернули такі слова як *беззахисна, оббитий* (тобто подвоєні приголосні не привертали уваги).

З експерименту ми побачили, що не носію мови дуже важко переказувати текст. Це пов'язано насамперед з тим, що більшість слів диктор не розумів. Саме переказ показав чи розуміє переповідач значення слів та взагалі зміст тексту. Це було помічено на прикладах, які були підібрані. Той, хто розумів значення, переказував текст, вимовляючи довгі та короткі приголосні і навіть було помічено, УЧ₁ давав пояснення [У судді було багато запитань, тобто суддя ставив багато запитань].

Довга вимова приголосних виділялась і давала більшу можливість пізнання голосу, ніж короткі. Кожен мовець звуки в тому самому слові промовляє по своєму, залежно від особливостей голосу, індивідуального тембру, манери говоріння тощо, про те фонемі усі сприймають однаково. Довгі фонемі сприймаються довго, короткі – коротко.

В нашому випадку розуміння тексту несли на собі **довгі** приголосні. Якщо науковці досліджували і досліджують лише голосні і говорять про їх вагомість, то наш експеримент вказує на **вагомість** приголосних, і не просто приголосних, а довгих та коротких приголосних, які несуть головну інформацію.

Як показав експеримент, особливості вимови довгих та коротких приголосних відіграють важливу роль у процесі ідентифікації мовця, та можуть стати одним із критеріїв фоноскопічного аналізу мовлення.

Література

1. Баранов А. Н. Введение в прикладную лингвистику. – М., 2000. – С. 8.
2. Бардіна Н. В. : Лінгвістична фоноскопія як новий напрямок прикладної лінгвістики (у друці).
3. Галяшина Е. И. Прикладные основы судебной фоноскопической экспертизы. //Теория и практика судебной экспертизы: Сборник . Питер, 2003. – С. 9–134.
4. Работягов А. В. Некоторые методические аспекты проведения судебной фоноскопической экспертизы. // Теория і практика судової експертизи і криміналістика. Зб., науково-криміналістичних матеріалів. Вип. 4. Харків „Право” 2004. – С. 219–227.
5. Русский язык и советское общество: Социол.-лингв. исследование / АНССР. Ин-т русс. яз.; Под. ред. М. В. Панова. – М.: Наука, 1968. – 213с.

В. Близнюк*

Київський національний лінгвістичний університет (Київ)
УДК 81'322

МОДЕЛЮВАННЯ ФОНЕМНОЇ СТРУКТУРИ ЛЕКСЕМ СУЧАСНОЇ ІТАЛІЙСЬКОЇ МОВИ

The analysis of phonemic structures of lexemes in a modern Italian literary language is carried out. The structure of a word is regard as a canonical form. Inductive and deductive linguistic model is a real-life and theoretically possible canonical forms are compared.

Модель (франц. *modèle*, італ. *modello*, від лат. *modulus* – міра, мірило, зразок, норма) в широкому значенні є умовним або мисленнєвим образом або прообразом якого-небудь об'єкта чи системи об'єктів («оригінала» даної моделі), що використовуються за даних умов в якості їхнього «заступника» або «представника» [Бульгіна, Крылов, 1990: 304].

Модель є лише формою вираження, яка може мати свою повну наукову значущість лише з урахуванням того, про що йде мова [3: 15].

Лінгвістичні моделі поділяють на індуктивні (аналітичні) й дедуктивні (синтетичні, породжувальні). Індуктивні моделі виходять з мовного матеріалу і шляхом його аналізу встановлюють класи й підкласи одиниць та їхню типологію. Дедуктивні моделі виходять з абстрактних класів, що є вихідними для моделі, і шляхом заданих операцій синтезу повинні породити, тобто обчислити, одиниці мови й визначити їхню ієрархію [4: 339].

Найпростішою моделлю фонемної структури лексем будь-якої мови є канонічна форма (КФ). Під каноном (від давньогрецького *kánon* – норма, правило, критерій) розуміється зведення положень, які мають догматичний характер, тобто є твердо встановленими й визнаними зразковими [6: 544]. Отже, канонічна форма є формулою, яка представляє слово в найбільш узагальненому вигляді – як ланцюжок послідовностей голосних (V) та приголосних (C) фонем [8: 284-286].

Предметом нашого дослідження є встановлення канонічних форм сучасної італійської літературної мови. Аналіз проведено на матеріалі механічної вибірки зі словника італійської мови «Dizionario Italiano Sabatini Coletti». – Firenze: Giunti Gruppo Editoriale, 1997, яка складає 20000 слів загальною кількістю 193688 фонем (83592 голосних і 110096 приголосних).

Завдання, що стоять перед нами, такі: аналіз розподілу довжини КФ; з'ясування моделюючої сили КФ певної довжини, встановлення КФ, які описують ядро, основну та периферійну систему лексичного складу сучасної італійської літературної мови; аналіз співвідношення теоретично можливих і реалізованих КФ, тобто зіставлення індуктивних і дедуктивних моделей КФ, і з'ясування причин, з яких можлива або неможлива реалізація тих або інших КФ.

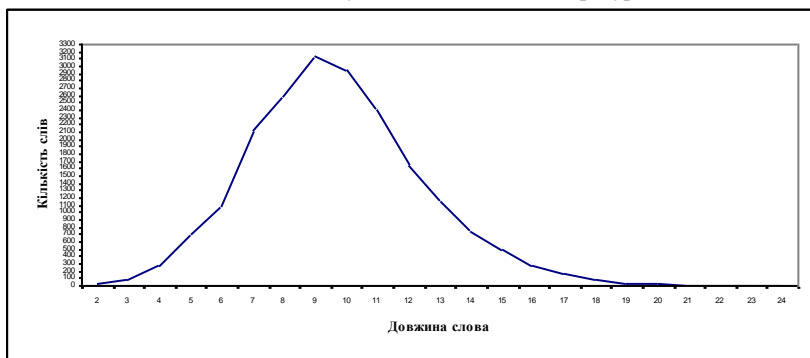
Перед тим, як починати дослідження правил побудови слів із фонем, необхідно встановити, які довжини слова допустимі в італійській мові. Аналіз розподілу довжини слів у вибірці показує, що найкоротшою є словоформа в 2 фонемі (в італійській мові є й словоформи довжиною в одну фонему, але їх дуже мало і вони не потрапили в нашу вибірку), а найдовшою є словоформа в 24 фонемі. До цього слід додати, що в словниках італійської мови зареєстровано найдовше слово в 26 фонем – /pre,ʃi:pi,tevo,lissi,mevol'mente/ [9: 65]. Найчастішими є слова довжиною дев'ять фонем. За ними слідує зі зменшенням частоти слова довжиною десять, вісім, одинадцять, сім фонем. Як впливає зі схеми 1, зростання кількості слів довжиною від двох до дев'яти фонем відбувається поступово, а зменшення кількості слів довжиною від десяти до двадцяти чотирьох фонем – більш інтенсивно. Отже, коливання середньої довжини слова в італійській мові становить від 9,64 до 9,72 ($\bar{x} \pm 2\sigma_{\bar{x}} = 9,64 \approx 9,72$).

Встановлений на матеріалі механічної вибірки інвентар КФ нараховує 1276 одиниць. Кожна КФ описує різну кількість слів. Дослідження моделюючої сили КФ показує, що в словах певної довжини виділяється лише одна КФ, яка має найвищу моделюючу силу. КФ з найвищою моделюючою силою описують 4816 слів або 24,08 % вибірки. Серед виявлених КФ форм можна виділити КФ з високою (від 40,85 % до 74,36 %), середньою (від 12,13 % до 43,48 %) та низькою (від 4,97 % до 9,73 %) моделюючою силою. 4 КФ (довжиною від 21 до 24 фонем) зустрічаються спорадично, тому для них не можна зробити статистично вірогідні висновки.

У словах довжиною від 2 до 7 фонем виділяється 1-3 моделі КФ, які описують до 75 % слів заданої довжини, або ядро системи лексики, та невелика кількість моделей, за якими будується до 90 % слів заданої довжини, або основна система лексики, а інші ж КФ, кількість яких є значною, належать до КФ з низькою моделюючою силою [5: 10]. Отже, для 2-фонемних слів сучасної італійської літературної мови 1 КФ (CV) описує 74,36 % слів такої довжини, для 3-фонемних 2 КФ (CVC, CVV) описують 67,61 %, для 4-фонемних 1 КФ (CVCV) – 68,23 %, для 5-фонемних 2 КФ (CVCCV, CCVCV) – 74,93 %, для 6-фонемних 2 КФ (CVCVCV, CCVCCV) – 65,15 %, для 7-фонемних 3 КФ (CVCCVCV, CVCVCCV, CCVCVCV) – 74,80 %.

Для слів довжиною від 8 до 14 фонем кількість КФ, яка необхідна для опису ядра і системи починає зростати, тому що зі збільшенням кількості фонем у слові зростає комбінаторна можливість сполучення фонем, і мова використовує більшу кількість КФ, які не суперечать її законам. Далі спостерігаємо зменшення кількості цих КФ, оскільки зменшується загальна кількість слів, що мають довжину від 14 до 24 фонем.

Схема 1. Розподіл довжини слів сучасної італійської літературної мови



Таблиця 1. КФ слів сучасної італійської літературної мови

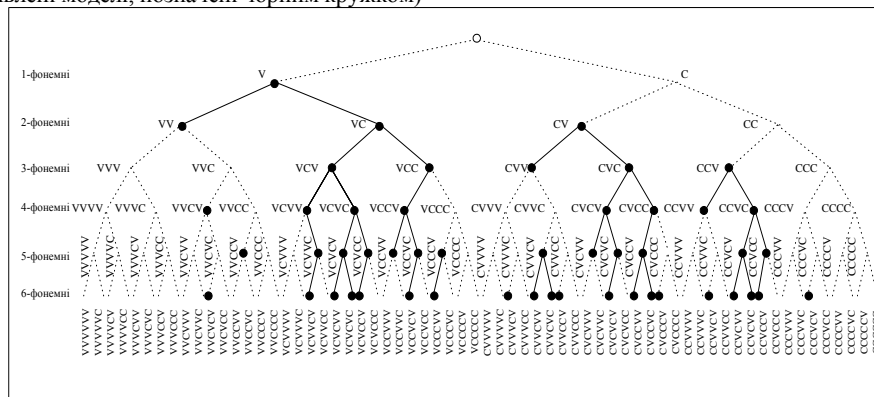
Довжина КФ	Кількість різн. КФ	Кількість слів даної довжини	КФ з найбільшою моделюючою силою			
			КФ	Кількість слів	Доля КФ в словах даної довжини, %	
2	3	39	CV	29	74,36	висока
3	5	71	CVC	29	40,85	середня
4	8	277	CVCV	189	68,23	висока
5	13	706	CVCCV	445	63,03	
6	20	1102	CVCVCV	555	50,36	
7	32	2123	CVCCVCV	923	43,48	середня
8	47	2591	CVCVCVCV	672	25,94	
9	66	3132	CVCCVVCV	803	25,64	
10	79	2953	CVCCVCCV	522	17,68	
11	115	2390	CVCCVVCVCV	290	12,13	низька
12	122	1644	CVCCVCCVVCV	160	9,73	
13	138	1146	CVCCVCCVCCV	84	7,33	
14	151	726	CVCCVVCVCVCV	41	5,65	
15	140	483	CVCCVVCVCVCVCV	24	4,97	
16	120	286	CVCCVVCVCVCVCVCV	19	6,64	
17	87	169	CVCVCVCVCVCVCVCV	15	8,88	
18	62	83	CVCCVVCVCVCVCVCVCV	5	6,02	
19	31	37	CVCVCVCVCVCVCVCVCV	3	8,11	
20	22	27	CVCCVVCVCVCVCVCVCVCV	4	14,81	середня
21	6	6	CVCCVVCVCVCVCVCVCVCVCV	1	–	спорадична
22	5	5	VCCVCCVVCVCVCVCVCVCV	1	–	
23	3	3	CVCVCVCVCVCVCVCVCVCVCV	1	–	
24	1	1	CVCCVCCVVCVCVCVCVCVCVCV	1	–	

Виділення КФ, що формують основну систему, включаючи ядро, можливе не лише в словах заданої довжини, але й цілком у вибірці. Залежно від моделюючої сили КФ створюється їхня ієрархія: 108 КФ моделюють 15157 слів (75,79 % всієї вибірки) і

становлять ядро основної системи. Основна система, включаючи ядро, охоплює 320 КФ, які моделюють 18174 слів (90,87 % вибірки). Останні 956 КФ моделюють 1826 слів (9,13 % вибірки). Такі КФ становлять периферійну підсистему.

Виходячи з результатів аналізу, можна порівняти кількість теоретично можливих і наявних моделей КФ слів заданої довжини в сучасній італійській літературній мові. Теоретично можливі різновиди фонемної структури слова можна легко встановити, побудувавши відповідне дерево, в якому кожне ліве відгалуження означає дописування голосної, а праве – приголосної до вже наявного ланцюжка фонем (фрагмент такого дерева наведено на схемі 2). Кількість таких структур для слів певної довжини дорівнює 2^n , де 2 – два класи фонем (V і C), а n – число фонем у слові. Отже, для однофонемних слів кількість різних КФ дорівнює $2^1 = 2$, для двофонемних – $2^2 = 4$, для трифонемних – $2^3 = 8$ і т. д. [4: 256].

Схема 2 Теоретично можливі та реалізовані моделі КФ слів сучасної італійської літературної мови довжиною від 2 до 6 фонем (примітка: суцільними лініями з'єднуються виявлені моделі, позначені чорним кружком)



Аналіз схеми 2 показує, що деякі КФ не можуть бути реалізовані внаслідок обмежень, які накладають закони мови: в італійській мові нема слів з структурою C, CC, CCC, CCCC і т. п. (за винятком декількох вигуків, що складаються з двох або трьох приголосних поспіль: /mm/, /ps/, /st/, /br/, /pr/, /pst/, які зареєстровані у словнику, але не потрапили до нашої вибірки). Не зустрічаються також сполучення чотирьох приголосних на початку слова [7: 66]. За результатами наших досліджень фонемної структури лексики сучасної італійської літературної мови збіг трьох голосних можливий лише в середині слова (за винятком двох вигуків /e'ɔ/, /u'o'ɛ/, які не потрапили до вибірки), тому не зустрічаються такі КФ, як VVV, VVVC, CVVV, а також КФ, в яких ми спостерігаємо чотири і більше голосних підряд: VVVV, VVVVC, CVVVV, VVVVV і т. п. Слід зауважити, що є такі КФ, які не суперечать законам мови, але не використовуються нею.

На підставі більш детального аналізу схеми 2 маємо такі результати.

Із 8 теоретично можливих КФ довжиною 3 фонем використовується 5, тобто 62,50%. Серед невикористаних – VVV, VVC, CCC. В чотирифонемних словах використовується 8 моделей КФ із 16 або 50,00 %. Не використовуються такі КФ: VVVV, VVVVC, VVVC, VCCC, CVVV, CVVC, CCCC, CCCC. Сполучення трьох голосних на початку чи в кінці слова або двох приголосних в кінці слова суперечать законам італійської мови. У п'ятифонемних словах використовується лише 13 моделей КФ із 32 теоретично можливих, тобто менше половини (40,63 %). Зростання частки нереалізованих КФ також пояснюється збільшенням кількості теоретично можливих моделей зі сполученнями трьох і більше голосних або приголосних на початку чи в кінці слова.

Частка використаних моделей КФ у словах довжиною шість фонем ще менша, в цьому випадку з 64 теоретично можливих знаходимо лише 20 КФ, тобто 31,25%.

Переважна більшість невикористаних КФ суперечить законам мови, тому що вони містять у собі сполучення трьох, чотирьох, п'яти і навіть шести голосних підряд, а також чотирьох, п'яти й шести приголосних.

Таблиця 2. Теоретично можливі та реалізовані моделі КФ слів сучасної італійської літературної мови

Довжина КФ	Кількість теоретично можливих КФ	Кількість реалізованих КФ	Процент реалізації КФ, %
2	4	3	75,00
3	8	5	62,50
4	16	8	50,00
5	32	13	40,63
6	64	20	31,25
7	128	32	25,00
8	256	47	18,36
9	512	66	12,89
10	1024	79	7,71
11	2048	115	5,62
12	4096	122	2,98
13	8192	138	1,68
14	16384	151	0,92
15	32768	140	0,43
16	65536	120	0,18
17	131072	87	0,07
18	262144	62	0,02
19	524288	31	0,01
20	1048576	22	0,002
21	2097152	6	0,0003
22	4194304	5	0,0001
23	8388608	3	0,00004
24	16777216	1	0,00001
Разом:	33554428	1276	0,004

Як випливає з таблиці 2, кількість різновидів теоретично можливих КФ форм дуже швидко росте зі збільшенням довжини слова. У той же час спостерігається зменшення процента реалізації цих форм у конкретних словах. Така картина пояснюється тим, що зі збільшенням довжини слова зменшується кількість довгих слів внаслідок обмежень сполучуваності двох класів (V і C) фонем. Отже, загальна кількість реалізованих КФ слів сучасної італійської літературної мови становить лише 0,004 % від теоретично можливих.

Висновки:

1. Опис лексем в термінах голосних (V) і приголосних (C) фонем дає можливість виявити КФ з найбільшою моделюючою силою, які описують значну кількість слів сучасної італійської літературної мови.
2. Невелика кількість КФ необхідна для опису ядра та основної системи слів заданої довжини, а також усієї лексики сучасної італійської літературної мови. Це є проявом закону переваги, за яким використовується лише невелика кількість одиниць мови, що формують мінімум, необхідний для її розуміння та запам'ятовування.
3. В сучасній італійській літературній мові, як і в усіх інших, використовується незначна кількість КФ з теоретично можливих. Це свідчить про великий потенціал невикористаних можливостей мови.

Література

1. Ахманова О. С. Словарь лингвистических терминов. – М., 1966. – 606 с.
2. Булыгина Т. В., Крылов С. А. Модель. // Лингвистический энциклопедический словарь. – М., 1990. – С. 304-305.
3. Лосев А. Ф. Введение в общую теорию языковых моделей. / Под ред. И. А. Василенко. Изд. 2-е, стереотипное. – М.: Едиториал УРСС, 2004. – 296 с.
4. Перебийніс В. С. Кількісні та якісні характеристики системи фонем сучасної української літературної мови: Дис. ... д-ра філол. наук. – К., 1970. – 510 с.
5. Перебийніс В. С. Модель // Українська мова: Енциклопедія. – К.: Укр. енцикл., 2000. – С. 339-340.

6. Слипченко Л. Д. Фонемная структура лексики английского языка: Автореф. дис. ... канд. филол. наук: 10.02.21. – К., 1974 – 21 с.
7. СЭС. Советский энциклопедический словарь. – М.: Сов. энциклопедия, 1982. – 1600 с.: ил.
8. Чельшева И. И., Черданцева Т. З. Итальянский язык // Языки мира: Романские языки. – М.: Academia, 2001. – С. 56-90.
9. Hockett Ch. F. A course in modern linguistics. – New York: The Macmillan Company, 1958. – 621 p.
10. Nespor M. Fonologia. – Bologna: il Mulino, 1994. – 348 p.
11. Sabatini F., Coletti V. Dizionario Italiano Sabatini Coletti. – Firenze: Giunti Gruppo Editoriale, 1997. – 3037 p. con illustrazioni

З. Дудник*
УДК 81'322.6

СЕГМЕНТАЦІЯ НА АКУСТИЧНІ СКЛАДИ ДЛЯ КОМПІЛЯТИВНОГО СИНТЕЗУ АНГЛІЙСЬКОГО МОВЛЕННЯ

This paper describes the approach to syllabification currently used in the Laboratory of Computational Linguistics (Kyiv National Linguistic University) for English text-to-speech synthesis. Syllable is a subcomponent of the word model whereas quasi-syllable is a subcomponent of the hierarchic speech model of phonetic word, rhythmical group, and syntagma. The real structure and syntactical position of quasi-syllables enables the TTS-system to retain the naturalness of concatenative speech signals.

1. Відомо, що в теорії складу залишається багато нез'ясованих питань, хоча над ними працюють тисячі дослідників. Тісно пов'язаними між собою є проблема складоподілу і процедура сегментування на склади з прикладною метою. Синтезатори мовлення «текст на озвучення», які підтримують конкатенацію складів, при доборі елементної бази спираються здебільшого на морфемно-фонологічні принципи складоподілу [1]. Процедуру препроцесорного фонетичного аналізу в цьому випадку прирівнюють до процедури автоматичного складоподілу графічного тексту, тобто джерелом правил силлабіфікації використовують матеріали, що фіксують результати мовного розвитку і функціонування. Натомість вимогою, яка висувається до конкатенованого мовлення на виході, є природність звучання, досягти якої можна лише з урахуванням реальних властивостей породження мовлення. Несумісність теоретичних підходів і практичних вимог вкотре ставить питання онтологічного підґрунтя теорії складу.

2. Виконання ручної сегментації акустичного сигналу мовлення на відрізки складової розмірності з метою створення акустичної бази складів в системі синтезатора англійського мовлення «текст на озвучення» привернуло увагу до можливостей розв'язання теоретичних проблем загальної фонетики і практичних проблем технології мовлення. Якщо за вихідну точку відліку обрати онтологічний аспект породження, то сегментація на склади має бути процедурою, що орієнтована на сутнісні властивості матерії мовлення, а саме на її психофізіологічно і артикуляторно обумовлену дискретність, а також на значну варіативність якості мовленнєвих подій. У такому разі широкі лінгвістичні узагальнення про склад з позиції виділення в ньому суттєвого, а також залежні від цієї позиції правила складоподілу мають братися до уваги з певною обережністю, адже їх онтологічна орієнтація – це результати мовного функціонування і нормативні вимоги, які не в змозі охопити реальні стратегії мовленнєвої поведінки.

Однак і для дослідника, який має справу з акустичним сигналом, вимовне зусилля розкривається через підручний матеріал, тобто є інструментально фіксованим результатом події. Тут так само виникає небезпека в тому, що особливості репрезентації мовлення за допомогою певного технічного прийому будуть ототожені із особливостями самого мовлення. Приміром, структуру акустичної події репрезентує часова розгортка сигналу (осцилограма, спектрограма), що відображує зміни тиску повітря та зміни передавальної функції мовленнєвого тракту, яку фіксує мікрофон із власними технічними характеристиками. Тому в акустичній картині немає повної відповідності артикуляційним зусиллям. Наприклад, коли язик сильно притискається до піднебіння і відштовхується від нього для наступного артикуляторного зусилля, акустичний сигнал як такий може бути відсутній, але це не означає, що немає артикуляційної події. Іншими словами, подія

* © З. Дудник, 2006