

на засадах стимулювання вітчизняних виробників і користувачів новітніми інформаційно-телекомунікаційними засобами і технологіями, комп'ютерними системами і мережами» [2].

Але вкрай несподіваним і незрозумілим є те, що у Доктрині наголошується, що удосконалення законодавства з питань ІБ і підготовка спеціалістів з питань ІБ є необхідними заходами держави виключно у воєнній сфері.

Крім того, у розробленій РНБО Доктрині зовсім відсутнє визначення терміну ІБ. В ній чітко не прописані основні цілі та задачі забезпечення ІБ в Україні. В Доктрині також не закріплена необхідність рішення на державному рівні актуальних проблем і задач в області ІБ, серед яких, наприклад, необхідність забезпечення управління (менеджменту) ІБ, управління інформаційними ризиками, аудиту систем ІБ, гармонізації національної нормативної бази у сфері ІБ та захисту інформації з міжнародною і європейською, а також рішення інших завдань ІБ на принципово новому (європейському) рівні. В прикінцевих положеннях Доктрини лише прописано, що «вона є основою для: формування державної політики у сфері інформаційної безпеки України; розроблення проектів концепцій, стратегій, цільових програм і планів дій із забезпечення інформаційної безпеки України; підготовки пропозицій щодо дальшого системного вдосконалення правового, методичного, науково-технічного і організаційного забезпечення інформаційної безпеки України» [2].

Таким чином, автори вважають, що хоча на даний час у національному законодавстві досить широко використовується термін «інформаційна безпека», але до кінця законодавчо так і не вирішена проблема його точного визначення, а також остаточно не розроблена цілісна система правового та нормативного забезпечення інформаційної безпеки України.

Література: 1. Соснін О.В. Проблеми державного управління системою національних інформаційних ресурсів з наукового потенціалу України : монографія / О.В. Соснін. – К. : Інститут держави і права ім. В.М. Корецького НАН України, 2003. – 572 с. 2. Доктрина інформаційної безпеки України. – Затверджена Указом Президента України від 8 липня 2009 року № 514/2009. 3. Біла книга Держспецзв'язу. – <http://www.dsszzi.gov.ua>. 4. Закон України «Про основні засади розвитку інформаційного суспільства в Україні на 2007-2015 роки» // Відомості Верховної Ради України, 2007, № 12, ст. 102. 5. Закон України «Про телекомунікації» // Відомості Верховної Ради України, 2004, № 12, ст. 155. 6. Конституція України // Відомості Верховної Ради України, 1996, № 30, ст. 141. 7. Закон України «Про Концепцію Національної програми інформатизації» // Відомості Верховної Ради України, 1998, № 27-28, ст. 182. 8. Закон України «Про основи національної безпеки України» // Відомості Верховної Ради України, 2003, № 39, ст. 351. 9. Указ Президента України «Про Воєнну доктрину України» від 15 червня 2004 року № 648/2004. 10. Закон України «Про захист інформації в інформаційно-телекомунікаційних системах» // Відомості Верховної Ради України, 2005, № 26, ст. 347. 11. Рішення Національної ради України з питань телебачення і радіомовлення від 1 липня 2004 року «Про забезпечення інформаційних потреб та інформаційної безпеки громадян». 12. Указ Президента України «Про рішення Ради національної безпеки і оборони України від 21 березня 2008 року “Про невідкладні заходи щодо забезпечення інформаційної безпеки України”» від 23 квітня 2008 року № 377/2008. 13. Концепція національної інформаційної політики : проект // Демократична Україна. – 2002. – 13 серпня.

УДК 681.391

АРТИКУЛЯЦІЙНІ ТАБЛИЦІ СЛІВ УКРАЇНСЬКОЇ МОВИ

Олена Архипова, Володимир Журавльов*, В'ячеслав Кумейко**

Національний технічний університет України „КПІ”, *Запорізький національний технічний університет, **Національний банк України

Анотація: Розглянуто задачі галузі дослідження якості каналів мовної комунікації. Складено десять артикуляційних таблиць слів української мови.

Summary: In this paper is considered tasks in the research field of the communications speech quality. Ten articulation tables of Ukrainian words are completed.

Ключові слова: Мовна інформація, розбірливість, артикуляційні таблиці.

І Вступ. Постановка задачі

Оцінка акустичних властивостей приміщення, каналів зв'язку та телекомунікаційних апаратів, а також захищеності мовної інформації пов'язана з необхідністю вимірювання і розрахунку розбірливості мови.

Усі відомі на цей час методи оцінки розбірливості мови можуть бути розділені на дві великі групи: *суб'єктивні експертні методи* і *об'єктивні методи* [1]. До суб'єктивних відносять такі методи, в яких складовою частиною вимірювальної системи є апарат мовлення та слуховий апарати людини, а до об'єктивних – методи, в яких весь процес вимірювання здійснюється приладами без участі органів чуття людини.

У наш час налічується кілька десятків об'єктивних методів розрахунку й вимірювання розбірливості мови. Ці методи можна поділити на формантні, теоретико-інформаційні, модуляційні та емпіричні [2]. Останнім часом впроваджено в практику низку нових комп'ютерних методів та сучасних апаратно-програмних вимірювальних комплексів. Об'єктивні методи вимірювання розбірливості мови досить перспективні внаслідок можливості значного прискорення й здешевлення процедури вимірювання. Проте рівень розвитку інструментальних методів поки що недостатній – досить проблематичним, наприклад, залишаються комплексне врахування факторів, що впливають на розбірливість (шуми, реверберації, луна, нелінійні спотворення тощо), оцінювання розбірливості мови у вокодерних цифрових каналах зв'язку та при наявності мовоподібних завад.

Артикуляційні випробування є суб'єктивним методом оцінки розбірливості – це найбільш прямий й очевидний, а іноді й єдиний спосіб дослідження якості каналу мовного зв'язку. Суть таких випробувань полягає в тому, що диктори зачитують (відтворюють) спеціально складену таблицю елементів мови (звуків, складів, слів або фраз), а аудиторі прослуховують ці елементи і роблять запис у відповідному бланку. Після закінчення циклу відтворення-прослуховування, який забезпечує достатній обсяг вимірювань, підраховується число правильно прийнятих аудиторіями елементів. Мірою розбірливості є відношення числа правильно прийнятих по каналу, що досліджується, елементів мови до загального числа переданих елементів, виражене у відсотках.

Головними перевагами методу артикуляційних випробувань є його універсальність, тобто можливість кількісної оцінки розбірливості для будь-якого типу каналу передачі мовних сигналів. Однак процедура організації суб'єктивних експертиз з оцінки розбірливості мови – справа громіздка, тривала й досить дорога.

На сьогодні в Україні відсутні національні стандарти в галузі артикуляційних випробувань, що враховують особливості звучання української мови. Замість них використовуються тимчасові документи типу Переліку [3], який дозволяє використовувати стандарти СРСР ГОСТ 16600-72 та ГОСТ 7153-85. Згідно з ГОСТ 16600-72 [4] розбірливість мови необхідно вимірювати з використанням російських артикуляційних таблиць звукосполучень та слів. Тип тестових таблиць істотно впливає на результати вимірювань. У зв'язку з відмінністю фонетичного складу мовних одиниць української мови від російської, ми вважаємо недоцільним використання вищевказаних (російських) артикуляційних таблиць для артикуляційних випробувань національних каналів передачі мовної інформації.

Для адекватного використання методу артикуляційних випробувань серйозною й самостійною науково-дослідною задачею є створення національних артикуляційних таблиць та стандартів.

II Основна частина

Артикуляційні таблиці складаються за певними правилами [5]. Ці правила враховують лінгвістичні (мовні) вимоги до таблиць. Лінгвістичні вимоги полягають у тому, щоб таблиці достатньою мірою відображали фонетичну структуру мови, співвідношення кількості звуків і букв, дотримували законів сполучення сусідніх звуків. Для забезпечення лінгвістичних вимог до артикуляційних таблиць для української мови необхідним є буквений та, у подальшому, фонемний частотні аналізи.

Частотний аналіз використання букв проведений для ряду європейських мов: французької, німецької, англійської, іспанської, італійської та російської [6]. Нами було проведено аналіз повторюваності букв української мови за допомогою програми, написаної мовою програмування C++ та пакету Excel [7]. У ході частотного аналізу оброблено біля 580 тисяч знаків (приблизно 100 тисяч слів) українських текстів технічного, художнього та публіцистичного спрямування. Отримано середні частоти появи букв українського алфавіту. Ці результати поруч із аналогічними даними для російської мови, зображені на рисунку 1 у вигляді порівняльної гістограми.

Для української мови ототожнено літери Г і Ґ, у зв'язку з тим, що в багатьох текстах вони не розрізняються, для російської мови ототожнено букви Е з Ё та Ъ з Ы.

Статистичний аналіз частот використання букв російської та української мов вказує на суттєве відхилення у розподілі букв, що свідчить про недоцільність використання результатів артикуляційних

випробувань на основі російських артикуляційних таблиць для випадків української мови і обґрунтовує актуальність поставленої задачі.

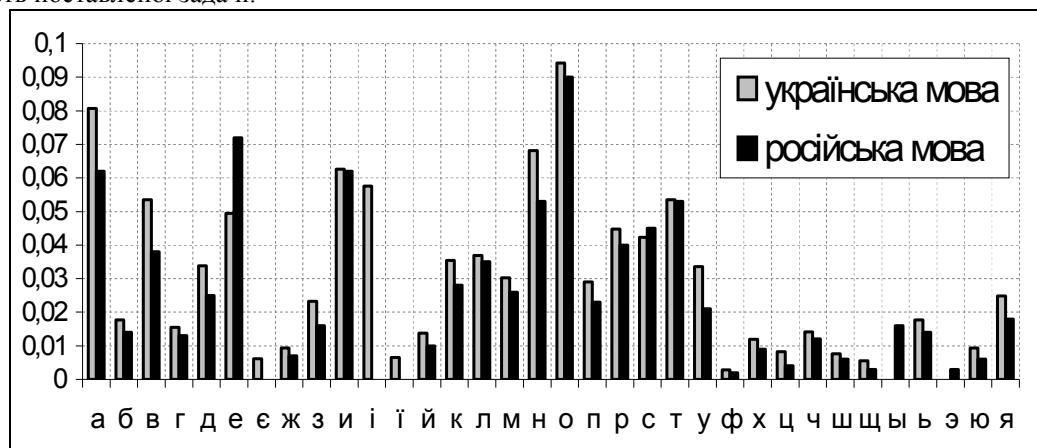


Рисунок 1 – Гістограма частот використання букв алфавіту української та російської мов

Враховуючи дані про частотний розподіл букв української мови (рис. 1), складено десять артикуляційних таблиць по 50 слів у кожній (таб. 1 – 10).

Таблиця 1

літо	п'ятсот	мова	квітка	човник	дідусь	чорниця	зовнішній	жахас	село
сало	позиція	скарга	пухнастий	пошта	невже	намисто	виправляти	змія	травичка
мідь	паління	прохолода	ідеї	гарбуз	сумки	усуди	марганцевий	щур	мазок
реве	буфет	дзига	сокира	витвір	гуси	ведмідь	теоретичний	льон	політика
обрії	коноплі	економ	знать	хашею	весна	обожноє	відновлення	тулуб	пісня

Таблиця 2

налаштувати	анімація	вперше	вагові	яма	морозиво	обробка	пізніше	прип'ятий	саджанець
тонкоброву	оксамит	нащадок	витяг	двері	засвоїти	контроль	загроза	осика	подужати
прослухаю	системно	узбіччя	м'ясо	відхід	документи	снідаю	лагодив	невірно	перев'язь
понівечений	ревнощі	солов'ї	фільтр	мінус	постійний	єдиною	ієрархія	тлумачник	тубус
белькотіти	гілля	вантами	судно	цикл	важливо	перстень	гличик	задній	келик

Таблиця 3

могузок	вхід	зібрати	прискіпливий	самітнику	торгаш	фундук	корелятор	дешевіє	дволикій
вишня	дещо	дівочий	по-осінньому	позаземне	постріл	ніхто	Закарпаття	сезонний	відбутися
доля	зате	ланцюга	недбальство	дохід	сором	сувої	вважаюче	бензин	українець
болісний	на	маховик	суперечливість	пиріжок	засць	новатор	симетрична	чагарник	вішун
вгадував	гніт	об'ява	кип'ятильник	муркіт	огорожі	плем'я	післямова	самолюб	антена

Таблиця 4

гусеня	канат	затишно	вдова	ямб	буча	позбутися	проти	коров'ячий	меліса
візок	злидні	розм'який	ослінчик	юніор	мовний	попекти	цибулина	спадкоємець	почуватись
закон	траса	всемогутній	циган	шуміти	жєврію	інтер'єр	беретися	нововведення	північ
щось	ледве	схвальний	зусилля	щедро	вакансії	самотньо	голодую	ультрафіолет	кіннота
м'ята	однина	плодоніжки	прихід	сім'ї	зуб	похмуріти	ховрашка	відверто	державна

Таблиця 5

захоплення	вогонь	напоїти	очманілий	ясенів	прохід	скандал	універсам	особову	якщо
красномовний	воли	листя	всуміш	перекіс	жебраку	покірно	виспівую	фотокамера	дія
зацікавленість	гудок	беркут	меншість	норов	підіб'єш	почути	розплескав	таємниця	через
змарнувала	велич	горище	засвітити	клімат	набутий	одежина	позолотить	домовини	юдей
трьохвимірний	юнаки	ягоди	відгуляв	дехто	збагнути	доцільної	нап'ясти	санчата	іржа

Таблиця 6

псевдонім	оцет	скеля	туман	заляканий	мамо	солому	лікарі	нейтраль	елегія
пополудні	риф	панцир	хаші	безволосий	вітаю	жіночка	їжачок	механізм	дуріть
підвестись	ноти	розпусно	тверезо	чотириріччя	абетка	мушля	гравець	оздоблює	книгарня
нагайка	туди	поверхово	програти	урівноваженість	вовна	насіння	дивосвіт	зморшки	німотою
від'їздити	б'єш	полагодив	пундик	якнайвище	свист	соболь	склянка	бідувати	слухати

Таблиця 7

скрегіт	традиції	одурманює	необачливо	пшенична	відлига	різьблений	хімія	хімеру	спорий
колесо	жовтень	продовольчі	мимоволі	дарований	піджак	претензія	гуща	телевізор	чутно
вибух	абиякий	неперервна	святковість	наступник	надлом	медицині	олово	казна-що	лунати
тютюн	ємність	знервованість	порваний	полегкість	слухання	кривобоко	вужі	парадокс	заміна
студія	багатії	успадковувати	безглуздя	вимикаю	фашист	золото	доця	ясночолі	миша

Таблиця 8

переплутаний	щавель	стандартно	хлопчик	мікроби	в'юни	корозія	кам'яних	буряк	обездолені
завбачливість	привілеї	журналіст	терміново	замахи	переніс	вигадка	грюкнути	чемно	солодкий
надокучливо	опівдні	поводитися	заметіль	совість	кутові	шурхіт	волосся	дієта	пасажирові
гайдамацтво	вбогість	полум'яно	звичайно	моціон	людоїд	равлик	сліпучий	щеня	бездоганно
експеримент	в'яжуть	шовінізм	абрикос	порція	завшир	стежина	секундант	нафта	аргументує

Таблиця 9

оптимально	вона	кардіолог	мілітарист	об'їзний	сичі	смерть	розчин	обмежив	шийка
автомобіль	угіддя	приватно	заплямований	комплекс	гурт	цибуля	вощений	пролягти	середні
передчуваю	наївні	балакаю	відпочинок	довголіття	борщ	кулька	напух	виноград	авжеж
роздавати	хтось	афішую	радіорозвідка	воєвода	заміс	зумисне	вовняну	нерішуче	маєтки
посилення	сіяти	стовбець	звукозахисні	помістити	несу	електрод	картопля	ціннісний	техніка

Таблиця 10

сивина	прикросі	ртуть	орієнтовний	консервна	акустику	тісто	давнина	заректисся	молот
найперше	типовий	вівса	епідемічний	підсмажити	удаватись	цегла	геофізик	дозвілля	згодом
купівля	небажане	цвіт	понаднормове	нестямною	снігуронька	волхв	відстані	безриб'я	щічка
з'ясувати	малятко	віче	підтоплення	поснідаючи	розігрівши	обоє	худоба	аналізатор	шкура
загальної	утюжок	цвях	мусульманство	переходити	бандероль	чохол	високої	рум'яний	килим

Представлені в таблицях українські слова не обмежені за кількістю букв і складів, належать до різних частин мови та надані у довільних відмінках. Ми вважаємо, що така форма відповідає реальній мові і не дозволяє передбачити закінчення слів, спотворюючи та завищуючи при цьому оцінку розбірливості.

У таблиці 11 надано значення частот появи букв та їх відхилення вибірково для другої і шостої таблиць слів.

Таблиця 11

	Частоти появи букв						
	середнє значення	Таблиця 2	відхилення	відхилення / середнє значення, %	Таблиця 6	відхилення	відхилення / середнє значення, %
а	0,0807	0,0774	0,00330	4,0892	0,0809	-0,00026	0,3255
б	0,0177	0,0172	0,00048	2,7240	0,0173	0,00033	1,8806
в	0,0535	0,0544	-0,00093	1,7368	0,0549	-0,00140	2,6189
г	0,0155	0,0172	-0,00168	10,8130	0,0173	-0,00183	11,7738
д	0,0338	0,0344	-0,00056	1,6449	0,0347	-0,00085	2,5262
е	0,0495	0,0516	-0,00208	4,2127	0,0462	0,00325	6,5633
є	0,0061	0,0057	0,00040	6,5170	0,0058	0,00035	5,7065
ж	0,0093	0,0086	0,00075	8,0321	0,0087	0,00068	7,2347
з	0,0232	0,0229	0,00030	1,3049	0,0231	0,00010	0,4491

Таблиця 11 (продовження)

	Частоти появи букв						
	середнє значення	Таблиця 2	відхилення	відхилення / середнє значення, %	Таблиця 6	відхилення	відхилення / середнє значення, %
и	0,0626	0,0630	-0,00041	0,6491	0,0607	0,00194	3,0929
і	0,0575	0,0602	-0,00267	4,6379	0,0549	0,00259	4,5067
ї	0,0065	0,0057	0,00074	11,4214	0,0058	0,00069	10,6534
й	0,0138	0,0143	-0,00050	3,6477	0,0145	-0,00063	4,5464
к	0,0354	0,0344	0,00100	2,8328	0,0376	-0,00219	6,1772
л	0,0369	0,0401	-0,00323	8,7524	0,0405	-0,00358	9,6953
м	0,0303	0,0287	0,00161	5,3231	0,0289	0,00136	4,5022
н	0,0681	0,0659	0,00220	3,2254	0,0665	0,00163	2,3863
о	0,0942	0,0917	0,00251	2,6682	0,0925	0,00172	1,8243
п	0,0290	0,0287	0,00039	1,3269	0,0289	0,00014	0,4713
р	0,0448	0,0458	-0,00107	2,4007	0,0462	-0,00147	3,2885
с	0,0424	0,0430	-0,00061	1,4324	0,0434	-0,00098	2,3119
т	0,0535	0,0544	-0,00096	1,7879	0,0549	-0,00143	2,6704
у	0,0336	0,0315	0,00204	6,0799	0,0318	0,00177	5,2656
ф	0,0028	0,0029	-0,00003	1,1114	0,0029	-0,00006	1,9881
х	0,0119	0,0115	0,00046	3,8751	0,0116	0,00036	3,0416
ц	0,0083	0,0086	-0,00032	3,8675	0,0087	-0,00039	4,7681
ч	0,0141	0,0143	-0,00018	1,2584	0,0145	-0,00030	2,1364
ш	0,0076	0,0086	-0,00100	13,1745	0,0087	-0,00108	14,1557
щ	0,0056	0,0057	-0,00016	2,8423	0,0058	-0,00021	3,7340
ь	0,0177	0,0172	0,00047	2,6667	0,0173	0,00032	1,8228
ю	0,0093	0,0086	0,00072	7,7541	0,0087	0,00065	6,9543
я	0,0248	0,0258	-0,00100	4,0161	0,0260	-0,00122	4,9180

Відносне відхилення від частоти використання букв українського алфавіту для всіх таблиць в середньому складає 4,8 %, що суттєво не впливає на результат артикуляційних випробувань. На даному етапі роботи застосування цих таблиць дозволяє суттєво підвищити точність оцінювання параметра якості каналів мовного зв'язку.

III Висновки

Сформульована актуальна науково-дослідна задача створення артикуляційних таблиць української мови.

Складено десять артикуляційних таблиць слів української мови з дотриманням частотного розподілу букв.

Планується доведення кількості артикуляційних таблиць до повного об'єму [1] та, за необхідності, складення таблиць із різним характером наборів слів (залежно від тематичного спрямування артикуляційних випробувань).

Література: 1. Покровский Н. Б. Расчет и измерение разборчивости речи [Текст] / Н. Б. Покровский. – М.: Связьиздат, 1962. – 392 с. 2. Дидковский В. С. Акустическая экспертиза каналов речевой коммуникации [Текст] / В. С. Дидковский, М. В. Дидковская, А. Н. Продеус. – К., 2008. – 420 с. 3. [Електронний ресурс] – Режим доступу: http://www.mintrans.gov.ua/mintrans/control/uk/publish/article?art_id=76171&cat_id=39278. 4. ГОСТ 16600-72. Передача речи по трактам радиотелефонной связи. Требования к разборчивости речи и методы артикуляционных измерений [Текст]. – М.: ИПК Издательство стандартов, 1973. – 90 с. 5. Вінницький В. П. Термінальне устаткування та передавання інформації в телекомунікаційних системах [Текст] / В. П. Вінницький, В. Г. Поліщук – К.: Політехніка, 2004. – 436 с. 6. Яглом А. М. Вероятность и информация [Текст] / А. М. Яглом, И. М. Яглом – М.: Наука, 1973. – 306 с. 7. Архипова О. О. Частотний аналіз використання букв української мови [Текст] / О. О. Архипова, В. М. Журавльов // Радіоелектроніка, інформатика, управління. – Запоріжжя, 2009. – № 2. – С. 53-56.