

УДК 811.411.21'374

## ВИКОРИСТАННЯ “ЧАСТОТНОГО СЛОВНИКА СУЧАСНОЇ АРАБСЬКОЇ МОВИ” У ВИКЛАДАННІ АРАБСЬКОЇ МОВИ

Богдан БАБАК

*Київський національний університет імені Тараса Шевченка, Інститут філології,  
бульв. Т. Шевченка, 14, кімн. 26, 01033, Київ, Україна,  
тел. (00380 44) 2393183*

Частотний словник може стати для викладача, дослідника мови джерелом інформації для прийняття необхідних рішень і корекції дій, що стали вже традиційними. Особливо актуальним частотний словник може стати для аматора читання оригінальних текстів, оскільки знання частовживаних слів стає очевидно необхідним, а засвоєння словникового запасу, що складається з рідковживаних слів, може додати процесу читання статус глибшого, насиченого фарбами, що допомагає відчутти мову автора.

*Ключові слова:* частотний словник, частотні списки, лінгвостатистичні закони.

Узимку 2006 року на відділенні сходознавства Київського національного університету імені Тараса Шевченка було завершено створення першого у світі “Частотного словника сучасної арабської мови”. Автором роботи „Частотний словник сучасної арабської мови” є студент 5-го курсу відділення сходознавства Бабак Б. Г., науковим керівником – професор Тищенко К. М., науковим консультантом – Рудий Б. А.

Проблематика цього дослідження полягає в тому, що частотні словники рідкісних мов залишаються здебільшого невідомими вітчизняному читачеві. Однак їхню значущість у викладанні мови, за умови правильного використання, дуже важко недооцінити.

Брак будь-яких публікацій чи досліджень щодо цього питання призвів до того, що автор створював “Частотний словник сучасної арабської мови” не спираючись на будь-які попередні частотні словники арабської мови, оскільки вони йому просто невідомі. Для власне арабських лінгвістів, як класичних, так і більшості сучасних характерним є “описовий” підхід у вивченні граматики мови. Той факт, що роботи з частотного дослідження арабської мови не проводили в арабському світі (чи принаймні невідомі авторові), укладач пов’язує з характерним менталітетом арабів.

Отже, цей словник є незалежним від впливів попередніх зразків. Роботи з його створення велися автономно, без будь-якого попереднього досвіду. З одного боку, це може позначитися негативно – велику кількість процесів та технологій для обробки інформації з урахуванням специфіки мови було запроваджено вперше. З іншого – укладачеві вдалося уникнути будь-якого копіювання помилок.

**Частотний словник** – вид словника (звичайно одномовного), у якому лексичні одиниці охарактеризовано з погляду їх вживання у сукупності текстів, характерних для мови загалом, або для окремого функціонального стилю чи для одного автора.

Залежно від типу лексичних одиниць можна виділити частотний словник словоформ, слів (лексем), основ слів (використовуються в інформатиці), слів у певних значеннях (семантичний частотний словник), словосполучень. Розрізняють абсолютні і відносні характеристики лексичної одиниці (x). Абсолютною характеристикою є частота (f) цієї лексичної одиниці (x), що дорівнює кількості вживань x у дослідженій сукупності текстів f(x). У частотному словнику подано або f(x), або нормовану частоту

$$\frac{f(x)}{N}$$

де N – кількість досліджених слів тексту. Відносною характеристикою вживаності лексичної одиниці є або її ранг (кількість лексичних одиниць, котрі в цьому частотному словнику мають абсолютну характеристику вживаності вищу чи однакову з абсолютною характеристикою певної лексичної одиниці), або будь-яка інша ознака, за якою ранг можна вирахувати з більшою чи меншою точністю. Переважно частотні словники представлені як абсолютні, так і відносні характеристики. Частотні словники використовують для створення ефективних методик викладання мови, для виділення ключових слів (в інформатиці), для створення раціональних кодів (у теорії зв'язку).

У цій статті автор намагається поетапно **обґрунтувати використання частотних словників у комплексному підході до вивчення іноземної мови**. Вивчення іноземних мов – праця нелегка, незалежно від того, чи самостійно вчиться людина, чи під керівництвом викладача. Останній варіант, безумовно, є ліпшим, оскільки викладач, навіть не будучи власне носієм мови, є носієм культури вивчення мов (або якої-небудь конкретної іноземної мови), котрий відображає величезний досвід попередніх поколінь, що ретельно вибирали найоптимальніший шлях для досягнення поставленої мети. Проте, сучасні технології здатні внести певні нововведення, оптимізуючи процеси запам'ятовування іноземних слів, читання оригінальних текстів, навчання відтворенню окремих звуків, слів, фраз і т. д.

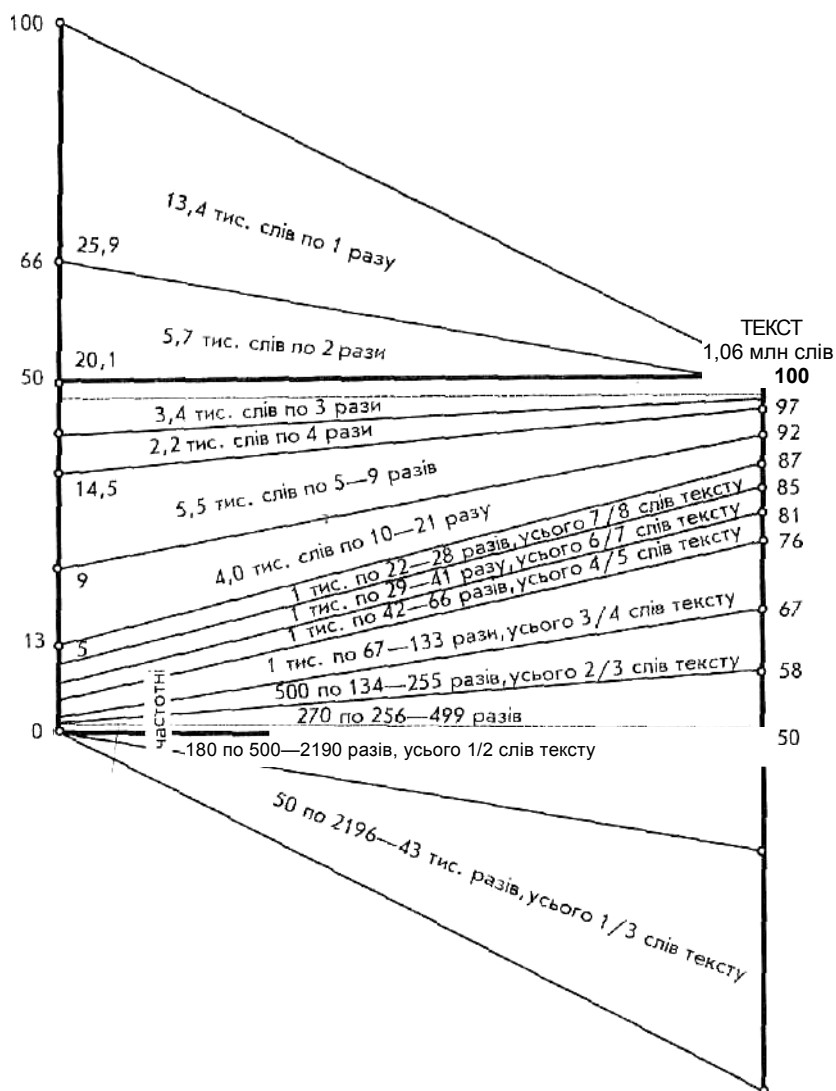
Таблиця 1

Лінгвостатистичні закономірності [1, с. 243]

№ Закономірності	ТИМ ВОНИ:	Формула
1. Есту–Зіпфа	вищі за рангом	$rf^r = \text{const}$
2. Зіпфа–Пірса	нечисленніші	$n = k_1 r^2$
3. Зіпфа–Флеша	коротші	$L = \sqrt{k_2 r}$
4. Зіпфа–Юла	багагозначніші	$Q = \sqrt[3]{f}$
5. Зіпфа–Гердана	багатовалентніші	$N_v = k_3 f$
6. Зіпфа–Арапова–Герц	давніші (старші)	$T = k_4 f$

Зі студентської лави ми пам'ятаємо – чим слово коротше, тим воно багатозначніше і тим частіше воно вживається в мовленні. Безумовно доцільним було б розпочинати навчання лексики іноземної мови саме з цих слів. Однак нерідко на заваді стає відсутність таких списків (словників) з рідкісних східних мов. Все ж, доклавши певних зусиль, їх можна знайти чи принаймі спробувати створити. Технології з їх створення є загально-доступними, що автор і спробує проілюструвати в межах цієї праці. Однак частотні списки у процесі вивчення мов часто безпідставно ігнорують.

Словник % 39,3 тис. слів



Співвідношення слів словника й тексту  
(схема К. М. Тищенка на підставі даних Л. М. Засоріної) [1, с. 243]

Чим же може допомогти частотний словник учителеві, учневі, дослідникові іноземної мови? Насамперед частотний словник, отриманий шляхом аналізу досить великого текстового фрагмента (новели, роману), демонструє багатство мови автора і його відмінність від словника “Елочки-людоджерки”. Наприклад, якщо ми візьмемося проаналізувати й порівняти словникові запаси Вальтера Скотта, Рафаеля Сабатіні (для нього англійська мова була не рідною, оскільки Р. Сабатіні за походженням італієць) і Джоанни Роулінг за їхніми творами “Айвенго”, “Одіссея капітана Блада” і “Гаррі Поттер і філософський камінь” відповідно, ми помітимо, що “Айвенго” містить 13 227 різних словоформ (слова можуть бути в різних формах, наприклад, *use, using, used* вважаємо різними словоформами), “Одіссея капітана Блада” – 9 498, а “Гаррі Поттер і філософський камінь” – лише 6 015, не дивлячись на фантастичну мову Хагрида (його мова істотно відрізняється від стандартної англійської). Про що свідчить ця статистика? Це свідчить про те, що людина зі словниковим запасом 2–4 тисячі слів без особливих проблем зможе читати “Гаррі Поттера” в оригіналі. З “Капітаном Бладом” справа буде трохи складніша, не кажучи вже про доблесного лицаря Айвенго.

Також викладач легко, використовуючи частотний словник, може визначити словниковий запас своїх учнів, необхідний для порівняно успішного читання тексту. При цьому, якщо текст досить великий, учневі (читачеві) цілком достатньо буде знань слів, що трапляються в тексті більше 2–3 разів, тоді як саме вони надають текстові того тематичного забарвлення, що є характерним тільки для певного автора.

Частотний словник також може вказати на особливості письмової мови автора аналізованих рядків. Якщо ми говоримо про текст англійською мовою, то в ньому, очевидно домінуючу позицію за частотою використання займатимуть артикль *the*, прийменники і займенники. Подібний висновок підтверджується вже поданим аналізом трьох вищезазначених добутків. Дійсно, лідерами є (за “Айвенго”):

Таблиця 2

Нумерація	Слово
1	The
2	Of
3	And
4	To
5	A
6	In
7	His
8	I
9	He
10	With

Приблизно аналогічну картину можемо спостерігати й на прикладі “Капітана Блада”. Але відмінною рисою “Гаррі Поттера” є наявність уже в першій десятці імені *Harry*. Також близько до вершини розташувалися імена *Ron, Hagrid*, і, трохи подалі, *Hermione*. Тут ми вже підходимо до функцій дослідника мови, оскільки часте використання в тексті імен може свідчити або про велику кількість діалогів, коли

герої звертаються один до одного на ім'я, або про оповідання в жанрі “action”, коли в тексті вдосталь виразів типу “Гаррі пішов”, “Рон зробив” і т. п.

Окрім того, частотний словник здатний допомогти дослідникові у визначенні частовживаних груп слів, що виділяють за якоюсь додатковою ознакою. Наприклад, нам необхідно довідатися: які визначення використали письменники для означення “недоброї” людини? Звичайний словник нам може дати приблизно такий список: *cad, git, heel, knave, rascal, rogue, scamp, scoundrel, villain* – негідник, мерзотник, сволота. Створюючи частотний словник, виявляємо всі нюанси вживання слів цієї групи в творах обраних авторів [2, с. 203]. Тут можна простежити залежність мови від часу життя, місця проживання, жанру, соціального статусу автора. Зокрема, слово “git” трапляється тільки в творах Дж. Роуллінг, слово “scoundrel” характерне лише для Р. Сабатіні, тоді як “knave” вжито в “Айвенго” 33 рази, у Р. Сабатіні – лише 2 рази, у “Гаррі Поттері” його зовсім нема.

Частотні словники (списки) по праву можуть зайняти чільне місце на ранніх етапах вивчення мови.

Продемонструємо яскравий приклад. Згадаємо „улюблені” підручники з англійської мови на ранніх етапах вивчення мови. З перших уроків нам пропонували вивчити „важливі” слова на кшталт *tooth-brush* чи *morning exercises*. Тепер на хвилинку замислимося і згадаймо скільки разів на день ми вживаємо словосполучення *tooth-brush* чи *morning exercises*. Результат очевидний і він дорівнює чи принаймі стрімко наближається до нуля. Натомість слова *I, he, she, it, to, from, under, with, in, on ...* в усному спілкуванні вживаються щохвилино.

Ще один приклад „*I go to school*”. Користуючись частотним словником англійської мови у процесі перекладу учень-початківець може не зрозуміти слів *go, school* – оскільки до першої десятки найуживаніших вони не входять. Слово ж *I* дасть йому чітке уявлення про суб'єкт дії – Я, а *to* про напрямок дії – туди. А це вже принаймні 50% розуміння висловлювання.

Автор не намагається якось принизити значущість слів на кшталт “атомохід” чи “втулка несучого гвинта”, все ж давати їх учням на ранніх етапах вивчення мови здається щонайменше ірраціональним. Така термінологія є безумовно важливою на етапі, коли базова лексика студента-перекладача вже сформована і постає потреба в розширенні певних „професійних” полів.

Програма, що генерує частотний словник, не повинна бути особливо складною, хоча бажане відсівання форм слова, що відрізняються від основної лише закінченням. Але, навіть не володіючи подібним інструментом, програма цілком може задовольнити потреби користувача, зацікавленого в одержанні загальної статистичної картини досліджуваного твору (або декількох творів). В Інтернеті автор знайшов два зразки генераторів. Один з них – повністю завершена “Програма-укладач частотних словників. Версія 2.6 Final” Володимира Пузанова. Інша – демо-версія “Частотный словарь. Версия 1.01” [4], що працює тільки з текстами істотно обмеженого обсягу.

Однак автор для укладання власного частотного словника використав програму Text Analyser, яку люб'язно надав її розробник Б. А. Рудий [http://www.langs.com.ua]. Програма дає змогу працювати з текстами з письмом справа наліво, а також з ієрогліфічним письмом.

Отже, частотний словник може стати для викладача, дослідника мови джерелом

інформації для прийняття необхідних рішень і корекції дій, що стали вже традиційними. Особливо актуальним частотний словник може стати для аматора читання оригінальних текстів, оскільки знання часто- вживаних слів стає очевидно необхідним, а засвоєння словникового запасу, що складається з рідковживаних слів, може додати процесу читання статус більш глибокого, насиченого фарбами, що допомагає відчувати мову автора.

Важливість частотного словника у вивченні іноземної мови цікаво проілюстрував Б. А. Рудий у його „Методі частотного штурму” [3].

**Сутність методу частотного штурму дуже проста:**

1. Ви берете текст, переклад якого Ви вже наперед знаєте.  
2. Отримуєте відранжований за частотністю частотний список слів цього тексту (за допомогою спеціальних програм).

3. Виділяєте у цьому списку верхівку з N-слів, яка покриває, наприклад, 50% слів тексту. Робите переклад цих слів на зрозумілу Вам мову. Завчаєте цю верхівку напам'ять.

4. Пробиєте читати вихідний текст, крокуючи відомими словами, а решту слів намагастесь зрозуміти, спираючись на те, що переклад речень Вам загалом вже відомий. (Звичайно, перед цим Ви мусите вивчити абетку та основні правила читання).

Так вдається швидко увійти в мову, — увійти у справжню, неадаптовану мову.

Чому це є проривом? Бо:

1. Ви виходите на рівень читання зі словником, а це значний етап.

2. Зусилля і час на це були максимально мінімальними.

Далі можна повторити процедуру з наступною порцією у частотному списку, щоб сукупна кількість відомих Вам слів покривала, наприклад, 80% слів тексту (але це вже буде не ~2% від усіх неоднакових слів, як у випадку з 50-відсотковим покриттям тексту, а ~20% усіх неоднакових слів). Читати після цього стане вже зовсім легко, бо 4-ри з 5-ти структурних елементів Вам будуть наперед відомі.

Важливо відзначити ще другий стовп методу частотного штурму: аналогічний ранжований список створюється для усіх типів структурних елементів мови (а не тільки для слів). Тобто, складається список: 1) афіксів (префіксів, суфіксів, закінчень); 2) граматичних правил (артиклі: які ознаки множини, роду, відмінків; іменники: які ознаки множини, роду, відмінків; займенники: у відмінках та основні типи; дієслова: якими є особові закінчення у теперішньому часі; якими є модальні дієслова та службові слова; прикметники: які ознаки множини, роду, відмінків; синтаксичні правила: порядок слів у оповідному, питальному та заперечному типах речення); 3) фонетичних правил (правила читання літер, правила вимови та інтонації); 4) фразеологізмів (основні важкоперекладні словосполучення на зразок „there is”, „do you”, і т. п.) та інших структурних елементів мови.

Так, можна зробити обґрунтовані висновки про те, що взявши “Частотний словник сучасної арабської мови” як посібник із засвоєння невідомої лексики на початкових етапах вивчення мови, можна суттєво полегшити викладання арабської мови у середній та вищій школах. Автор також залишає за собою право розробляти подальші методичні посібники, використовуючи “Частотний словник сучасної арабської мови”.

Подана нижче перша сторінка частотного списку „Частотного словника сучасної арабської мови” може дати уявлення про отримані результати:

Таблиця 3

Нумерація	Слово	F
1.	و	5626
2.	في	3398
3.	من	2513
4.	أن	1468
5.	على	1448
6.	إلى	908
7.	الله	893
8.	عن	789
9.	لا	774
10.	التي	675
11.	أو	623
12.	ما	622
13.	كان	553
14.	هذا	524
15.	قال	498
16.	الذي	489
17.	هو	481
18.	ذلك	477
19.	هذه	456
20.	مع	440
21.	إن	419
22.	عليه	384
23.	لم	378
24.	قد	354
25.	له	350
26.	هي	293
27.	من	292
28.	كل	272

Актуальність цього дослідження визначається також цілковитою відсутністю на теренах України будь-яких праць, як наукових, так і практичних, присвячених висвітленню теми арабських частотних словників. Роботи з дослідження саме арабської мови в частотному аспекті на теренах України не ведуться. Цією працею маємо на меті запровадити такі дослідження на науковій основі і, відповідно, полегшити процес вивчення самої мови.

Залишаємо за собою право продовжувати розробки проектів на основі “Частотного словника сучасної арабської мови”. Можливе також створення комплексного підручника з арабської мови для початкових рівнів навчання за технологією частотності з застосуванням автентичних текстів попередньо

адаптованих до читання. А також розташування граматичних тем у навчальному процесі згідно з їхнім рангом частотності (у попередньо складених частотних списках), що, на нашу думку, позитивно вплине на результативність процесу викладання.

Варто також зауважити, що необхідні частотні списки можна укласти і для окремих частин мови. Однак, не маючи змоги дістати необхідну кількість друкованого матеріалу, ми використовували електронні варіанти з мережі Інтернет. Тому за об'єкт дослідження взято сучасну літературну арабську мову Інтернету, що своєю чергою позитивно позначилося на актуальності отриманих результатів.

1. *Тищенко К. М.* Метатеорія мовознавства / К. М. Тищенко. – К., Основи, 2000.
2. *Eisenson H.* Scanners And Secret Frequencies / Henry Eisenson. (Electronic Underground), Paladin Press (January 1994).
3. <http://www.langs.com.ua>
4. <http://www.rowergum.narod.ru>

## USAGE OF “THE FREQUENCY DICTIONARY OF MODERN ARABIC” IN THE TEACHING THE LANGUAGE

**Bogdan BABAK**

*Taras Shevchenko Kyiv National University, Institute of Philology,  
T. Shevchenko av., 14, room 26, 01033, Kyiv, Ukraine,  
tel.: (0038 044) 2393183*

A frequency dictionary can be for the teacher, language researcher the source of the information needed, the generator for making necessary decisions and correction of actions which have already become traditional. A frequency dictionary can become of vital importance for the amateur-reader of original texts, as the familiarity with frequently used words obviously becomes a necessity, and mastering of vocabulary which consists of rarely common words can add the process of reading status of more deep, saturated paints, allows to feel the language of author.

*Key words:* Frequency dictionary, frequency list, linguistically laws.

Стаття надійшла до редколегії 27.09.2007

Прийнята до друку 03.03.2008