

UDC 811.581

QUANTITATIVE CHARACTERISTICS OF PRIMARY LEVEL CHINESE TEXTBOOKS

N. Isaieva

DSc (Philology), Associate Professor
Educational and Scientific Institute of Philology
Taras Shevchenko National University of Kyiv
14, Tarasa Shevchenka Blvd., Kyiv, 01601, Ukraine
inattalia@gmail.com
ORCID: 0000-0003-3282-4752

I. Sytnyk

PhD Student (Philology)
Educational and Scientific Institute of Philology
Taras Shevchenko National University of Kyiv
14, Tarasa Shevchenka Blvd., Kyiv, 01601, Ukraine
sytnykilona@gmail.com
ORCID: 0000-0002-9307-6180

Despite the fact that textbooks are central in Chinese language learning, only limited research has explored to what extent textbooks' content is relevant from a vocabulary perspective. In recent years much work on the potential of quantitative database of Chinese elementary textbooks has been carried out in modern Chinese linguistics; however there is still considerable disagreement with regard to practical objective data on vocabulary used in Chinese language textbooks at different levels. This paper focuses specifically on the quantitative analysis and aims to introduce the quantitative characteristics of vocabulary in four Chinese elementary textbooks: "Practical

© 2022 N. Isaieva and I. Sytnyk; Published by the A. Yu. Krymskyi Institute of Oriental Studies, NAS of Ukraine on behalf of *The Oriental Studies*. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by-nc/4.0/>).

Chinese Reader” “新实用汉语课本”, “Developing Chinese. Comprehensive course. Beginner level” “发展汉语初级综合”, “Boya Chinese. Elementary” “博雅汉语初级”, “Chinese course” “汉语教程”. In order to identify quantitative database the corpus has been constructed based on the words in the textbooks. The main stages of building text corpus of Chinese elementary textbooks are outlined. Statistical significance was analyzed through the use of linguistic Internet program “Chinese Corpus online”. The research results are presented. Based on the database, statistical analyses, such as the index of vocabulary diversity, the index of exclusivity, the average repetition of the word in the text, etc., are conducted. This work has proved that there is great variation in the number and selection of words across the books indicating that there does not seem to be a common thought behind word selection in textbooks used to teach Chinese as a foreign language. The present findings have important implications for future textbook development and may be used for writing Chinese textbook for Ukrainian students. Future studies on the current topic are therefore required in order to establish a frequency dictionary and a quantitative proficiency guideline for material development.

Key words: Chinese language, statistical analysis, quantitative characteristics, textbook, elementary level

Н. С. Ісаєва, І. В. Ситник

КВАНТИТАТИВНІ ХАРАКТЕРИСТИКИ ЛЕКСИЧНОГО СКЛАДУ ПІДРУЧНИКІВ З КИТАЙСЬКОЇ МОВИ ПОЧАТКОВОГО РІВНЯ

У китайській лінгвістиці ХХ–ХХІ століть набули особливої актуальності лінгвостатистичні методи дослідження мови, що дають змогу дістати кількісну характеристику тексту, яка доповнює або стає підґрунтям якісного лінгвістичного аналізу. Варто зазначити, що квантифікація мови тісно пов’язана з будь-якою її категоризацією. Наприклад, китайська мова характеризується певною кількістю складоморфем, словник мови характеризується певною кількістю лексем, які формують лексико-семантичні поля за своїми кількісними характеристиками. Особливо продуктивним, на наш погляд, є використання лінгвостатистичних методів у дослідженні педагогічного дискурсу китайської мови, оскільки кількісні показники лексичного складу навчальних текстів є вкрай важливими для створення навчально-методичних матеріалів (підручників, посібників,

збірників текстів для домашнього читання тощо), які б задовольняли потреби сучасної комунікації.

У процесі вивчення лексики підручників з китайської мови китайські лінгвісти зазвичай визначають регулярність розподілу лексичних одиниць у текстах та їхню статистичну структуру. Існує вже певна традиція використання лінгвостатистичного методу для виявлення особливостей лексичного складу підручників з китайської мови [褚琦 2018; 江子莹 2019; 王淼 2017; 王凤玲 2020; 孙彦 2019]. Деякі лінгвостатистичні дослідження виконано на матеріалі підручників з китайської мови початкового рівня, зокрема праця Сунь Яня (孙彦) «Дослідження лексичного складу підручника “Новий практичний курс китайської мови” (початковий рівень)» [孙彦 2019] містить інформацію про кількість вживань кожної лексеми в текстах і вправах підручника. Укладено частотний словник підручників “Курс китайської мови” [黄润杉 2018], “Поглиблений курс китайської мови” [耿直 2019; 刘诗涵 2016; 周晓晔 2020], “Комплексний курс китайської мови *Фачжань ханьюй*: початковий рівень” [江子莹 2019; 李小玲 2019] тощо. З огляду на обсяг і різноманіття результатів цих досліджень постає необхідність компаративного аналізу лінгвостатистичних показників наявних досліджень з метою виявлення й опису особливостей спільного лексичного складу (зокрема, найчастотнішого) підручників для подальшого опрацювання з прагматичною метою (укладання нових навчальних матеріалів, орієнтованих на потреби українських студентів). Тож актуальним видається й комплексний лінгвостатистичний аналіз лексичного складу зазначених підручників з китайської мови початкового рівня. Проте порівняння в кількісному ракурсі лексичного складу текстів зазначених підручників здійснюється вперше.

Мета розвідки – дати кількісно-статистичну інтерпретацію лексики китайськомовного педагогічного дискурсу початкового рівня та виявити характерні особливості розподілу лексем у текстах підручників.

Мета статті зумовила необхідність розв’язання низки завдань: проаналізувати останні дослідження лексичного складу підручників з китайської мови початкового рівня; укласти частотний список лексичних одиниць текстів підручників; виявити

статистичні параметри лексичного складу текстів, зокрема величину покриття тексту для певного рангу, обсяг словника, індекс словникової різноманітності, середню повторюваність слів, індекси винятковості та концентрації.

Для досягнення поставленої мети були використані взаємодоповнювальні методи лінгвістичного аналізу: компаративний метод – для дослідження лінгвостатистичних характеристик низки підручників з китайської мови початкового рівня, представлених у працях сучасних китайських лінгвістів; методи корпусної лінгвістики – для створення значного за обсягом, представленого в електронному вигляді, уніфікованого і структурованого корпусу текстів як основи для аналізу; статистичний аналіз – для підрахунку відсоткового співвідношення кількості лексем, укладання частотного списку і виявлення частотних закономірностей лексичного складу підручників за допомогою статистичних формул; кореляційний аналіз – для порівняльного аналізу частот лексем у досліджуваному матеріалі.

Джерельною базою квантитативного дослідження китайськомовного педагогічного дискурсу є статистичні дані корпусу текстів і діалогів (стилістично однорідні) підручників з китайської мови початкового рівня, виданих у Китаї (дотримання просторової єдності) з 2006 до 2011 року (досить вузький часовий зріз), а саме: “Новий практичний курс китайської мови” “新实用汉语课本” [刘珣 2002], “Комплексний курс китайської мови *Фачжсань ханьюй*: початковий рівень” “发展汉语初级综合” [荣继华 2011], “Поглиблений курс китайської мови. Початковий рівень” “博雅汉语初级” [任雪梅, 徐晶 2004], “Курс китайської мови” (початковий рівень) (“汉语教程”) [杨寄洲 2009]. Увага до цих підручників пояснюється тим, що сьогодні їх активно використовують на практичних заняттях з китайської мови у всьому світі, а також під час підготовки до міжнародного іспиту з китайської мови HSK [郭曙纶, 杨晓惠 2016]. При створенні корпусу було дотримано змістових принципів відбору текстів власне навчальних підручників [Карамнов 2013]: принцип цілісності (навчальні підручники з китайської мови початкового рівня написані згідно з вимогами навчальної програми “Рівні володіння лексикою та ієрогліфами сучасної китайської мови” (кит. “汉语水平词汇与汉字

等级大纲”) [国家汉语... 2001]), принцип відповідності проблемній ділянці (корпус текстів і діалогів для читання), принцип структурної спрямованості (корпус основного тексту підручника з китайської мови).

“Новий практичний курс китайської мови” “新实用汉语课本” (далі НПККМ) – це підручник з китайської мови як іноземної, виданий на початку ХХІ ст. авторським колективом Пекінського лінгвістичного університету (北京语言大学) за редакцією Лю Сюня (刘珣), видатного професора в галузі викладання китайської мови як іноземної. Цей комплект підручників складається із шести томів. З першого по четвертий том – це початковий рівень, а п’ятий та шостий – середній рівень.

Дослідженням НПККМ займалися Цао Маньвень (曹漫雯), Чжан Ніннін (张宁宁), Сунь Янь (孙彦), Чу Ці (褚琦). Зокрема, Цао Маньвень (曹漫雯) здійснила порівняльний аналіз двох комплектів написаних і виданих у Китаї підручників з китайської та англійської мов, а саме НПККМ та “Нова концепція англійської мови” (“新概念英语”) [曹漫雯 2018]. З’ясовуючи теоретичні проблеми, що виникають при формуванні словникового складу підручників з китайської мови як іноземної, Цао Маньвень (曹漫雯) розмірковує над тим, як можуть бути використані ефективні методи навчання підручника з англійської мови для китайців “Нова концепція англійської мови” при написанні підручників з китайської мови для іноземних студентів. Чжан Ніннін (张宁宁) у своїй роботі порівняв НПККМ з підручником “Поглиблений курс китайської мови. Початковий рівень” “博雅汉语初级” [张宁宁 2015]. На підставі порівняльного аналізу двох комплектів підручників за всіма аспектами (граматика, словниковий склад, вправи тощо) науковець визначає способи покращення якості написання майбутніх підручників з китайської мови. Сунь Янь (孙彦) у своїй статті [孙彦 2019] досліджує частотність вживання ієрогліфів у НПККМ початкового рівня. Зокрема, він визначив 10 найчастотніших ієрогліфів: “我、你、是、不、他、学、的、们、有、语”. Чу Ці (褚琦) порівнює НПККМ з підручником “Комплексний курс китайської мови *Фачжань ханьюй*: початковий рівень” “发展汉语初级综合” (далі ФХ) в різних аспектах [褚琦 2018]. Наведемо лише кілька важливих висновків Чу Ці

(褚琦). По-перше, у словниковому складі аналізованих підручників простежується суттєва різниця. Загальна кількість слів у НПККМ (I, II) становить 1569, а в ФХ – 1161 слово. Крім того, варто зазначити, що кількість слів у I та II томах НПККМ становить 867 та 486, а в I та II томах ФХ – 702 та 675. Загальна кількість нових слів у кожному уроці становить 29 слів у ФХ та 45 в НПККМ. Отже, праця Чу Ці (褚琦) засвідчує відсутність узгодженості у відборі лексики при написанні підручників з китайської мови початкового рівня.

“Комплексний курс китайської мови *Фачжсань ханьюй*: початковий рівень” “发展汉语初级综合” – підручник, виданий Пекінським лінгвістичним університетом (北京语言大学). У вступі до цього підручника згадується, що весь комплект підручників написаний згідно з принципом “інтеграція, розмаїття та інновація”. Комплект підручників ФХ поділений на три рівні (початковий, середній та високий) і п’ять серій (комплексний курс, розмовна мова, аудіювання, читання, письмо). Без сумніву, саме комплексний курс є синтезом усіх інших чотирьох серій та сприяє комплексному розвитку всіх навичок, натомість інші чотири серії спрямовані на одну навичку, а саме розмовну мову, або аудіювання, або читання, або письмо.

Питання переваг та недоліків підручника ФХ в китайських наукових колах порушувалося в працях Цзян Цзиїна (江子莹), Лі Сяоліна (李小玲), Лю Їна (刘颖). Зокрема, Цзян Цзиїн (江子莹) досліджував методи ефективного навчання китайської ієрогліфіки [江子莹 2019]. Ознайомлення з культурою Китаю як тематичним складником підручників з китайської мови було предметом дослідження Лі Сяоліна (李小玲) [李小玲 2019]. Лю Їн (刘颖) зробив внесок у дослідження проблеми репрезентації та частоти повторення нової лексики в підручниках. Дослідник розглядав питання лексичного складу підручників на початковому етапі, а також інші питання лексики китайської мови теоретичного і практичного характеру [刘颖 2016]. Усі згадані праці стосуються окремих аспектів ФХ, однак цілісні частотні словники китайської мови на різних її рівнях є актуальним напрямком, який залишається на стадії незавершеності.

“Поглиблений курс китайської мови. Початковий рівень” “博雅汉语初级” (далі ПККМ) – підручник з китайської мови початкового рівня, розроблений за редакцією професорки Пекінського університету (北京大学) Лі Сяоці (李晓琪) у 2004 р. Стрімке зростання популярності ПККМ зумовило те, що у 2013 р. видавництво Пекінського університету (北京大学出版社) перевидає вже якісно новий варіант підручника. Комплект підручників ПККМ поділяється на три рівні (початковий, середній, високий). Початковий рівень ПККМ складається із двох томів. У першому томі 30 уроків, а в другому – 25.

З моменту видання підручник ПККМ викликає зацікавлення науковців. Ретельний аналіз останніх розвідок окреслив два основні напрями наукових досліджень ПККМ, а саме визначення особливостей самого підручника ПККМ або порівняння основних характеристик ПККМ з іншими підручниками. Дослідженням суто особливостей ПККМ займалися Ген Чжи (耿直) [耿直 2019], Чжоу Сяо (周晓晔) [周晓晔 2020], Бай Ваньїн (白婉莹) [白婉莹 2020] та інші. З метою виявлення недоліків та переваг “博雅” Чи Хуейцянь (迟惠倩) [迟惠倩 2017], Хоу Яньдін (侯艳婷) [侯艳婷 2018], Чжао Даньхуей (赵丹慧) [赵丹慧 2015] та інші зіставляли ПККМ з іншими підручниками з китайської мови. Особливу увагу дослідників привертає кількісний склад лексики підручника.

Ген Чжи (耿直), аналізуючи лексику початкового рівня ПККМ, встановив, що: 1) у корпусі текстів, укладеному на основі 55 уроків, довжиною 9364 слова зафіксовано 1658 різних слів; 2) з 1658 слів 793 слова зустрічаються лише один раз, що становить майже половину (47,83 %) усієї кількості слів; 3) у 65 уроках підручника “Курс китайської мови” (початковий рівень) (“汉语教程”, далі ККМ) налічується 1328 слів, що менше, ніж у ПККМ; 4) хоча в підручнику ККМ кількість слів менша, ніж у ПККМ, але частотність їхнього вживання в підручнику ККМ значно перевищує ПККМ (15599 та 9364 відповідно) [耿直 2019]. Отже, кількість слів у підручниках та частотність вживання значно відрізняються, що потребує подальших досліджень проблеми відбору лексичного складу підручників.

“Курс китайської мови” (початковий рівень) (“汉语教程”) – це підручник з китайської мови, написаний за редакцією Ян Цзічжоу (杨寄洲) та виданий Пекінським лінгвістичним університетом (北京语言大学). Після видання підручник був високо оцінений та обраний для викладання китайської мови багатьма університетами в КНР та за кордоном. Підручник призначений для навчання початківців і складається з трьох томів, кожен з яких ділиться на дві частини (上/下). Загалом у підручнику 76 уроків, які студенти мають опрацювати за один навчальний рік. Перший том складається з 30 уроків, другий том – із 20-ти, третій – із 26-ти. Автори підручника зазначають, що словниковий склад підручника становить 2800 слів. Ґрунтовні дослідження особливостей підручника ККМ здійснили Хуан Жуньшан (黄润杉) [黄润杉 2018], Ван Фенлін (王凤玲) [王凤玲 2020] і Ван Мяо (王淼) [王淼 2017].

Отже, всі вищезазначені дослідження можна розділити на дві групи: перша група – це комплексні дослідження різних комплектів підручників, наприклад НППКМ, ФХ і ПККМ, з порівняльним аналізом різних аспектів, як-от: структура, граматичні конструкції, тематичне наповнення, словниковий склад тощо; друга група охоплює порівняльний аналіз комплектів двох різних підручників за однією характеристикою, наприклад структурою. Дослідники висловлюють різні погляди, керуючись однією метою, яка полягає в дослідженні та удосконаленні навчальних матеріалів з китайської мови як іноземної.

Представлене дослідження є першою спробою у вітчизняній синології за допомогою статистичного методу здійснити аналіз лексичного складу текстів китайських підручників початкового рівня “Новий практичний курс китайської мови” “新实用汉语课本”, “Комплексний курс китайської мови *Фачжань ханьюй*: початковий рівень” “发展汉语初级综合”, “Поглиблений курс китайської мови. Початковий рівень” “博雅汉语初级”, “Курс китайської мови” (початковий рівень) (“汉语教程”). Одержані квантитативні дані уможливають подальше визначення шару найчастотнішої лексики китайської мови, представленої в межах зазначених підручників.

Статистичний аналіз лексичного складу текстів китайських підручників початкового рівня передбачає виконання послідовності певних етапів:

1) відцифрування текстів;

2) вичитування та вилученням вказівок на імена дійових осіб (丁力波, 陆雨平, 林娜, 力波, 陈老师, 马大为, 宋华, 大为, 王小云, 杨老师, 张教授, 丁云). Тексти підручників подаються у формі діалогу між двома або більше дійовими особами. Аналізу піддалася власне мова персонажів;

3) цифри, передані знаками від 0 до 9, розписувалися ієрогліфами та рахувалися як одне слово. Наприклад, 55 ієрогліфами записується 五十五 та рахується як одне слово;

4) вилучення слів, написаних латиницею, наприклад ОК, e-mail;

5) видалення всіх знаків пунктуації (, 。 ? ! “”、 《》 ; :), тому що програма рахувала їх як окреме слово;

6) усунення омонімії шляхом додавання до одного з омонімічної пари умовної позначки. У такий спосіб ці слова графічно розрізнялися, і програма рахувала частоту вживання кожного з них окремо;

7) формування корпусу текстів;

8) у комп'ютерній програмі Chinese Corpus online – Frequency statistics of words (字词频率统计) [语料库在线... 2004] одиницею автоматичного підрахунку можна обрати ієрогліф (字) або слово (词), на цьому етапі статистичного аналізу одиницею автоматичного підрахунку обрано слово (词);

9) автоматичний підрахунок кількості слів (词频统计);

10) автоматичний обрахунок абсолютної та відносної частоти кожного слова корпусу за допомогою комп'ютерної програми [语料库在线... 2004]. Абсолютна частота події в лінгвостатистиці – це кількість реалізацій цієї події в процесі дослідження. Відносна частота події – це відношення абсолютної частоти до числа проведених дослідів або до числа одиниць у дослідженому масиві;

11) розташування всіх слів шляхом сортування в порядку зменшення частоти та надання кожному слову рангу, тобто порядкового номера слова в цьому списку. Результатом цієї операції

є ранговий словник слів НПКМ, ФХ, ПМКМ та КМ із вказівкою на абсолютну та відносну частоту використання слів.

У результаті проведеної роботи було отримано частотний список слів НПКМ, ФХ, ПМКМ та КМ, які посортовано за спадом частот. У корпусі текстів НПКМ довжиною 22948 слів (N) зафіксовано 2764 різних слова (V). У корпусі текстів ФХ довжиною 13871 слово (N) зафіксовано 2149 різних слів (V). У корпусі текстів ПМКМ довжиною 8675 слів (N) зафіксовано 1819 різних слів (V). У корпусі текстів КМ довжиною 22323 слова (N) зафіксовано 3767 різних слів (V).

На основі отриманих частотних даних розглянемо виявлені лінгвостатистичні закономірності лексичного складу китайських підручників початкового рівня. Програма Chinese Corpus online – Frequency statistics of words (字词频率统计) [语料库在线... 2004] автоматично обрахувала абсолютну та відносну частоту всіх слів. У частотному списку слова розміщено в порядку спадання частот: слово з найбільшою частотою має ранг 1, наступне – 2 і т. д. Отриманий ранговий список слів дає змогу обчислити величину покриття тексту для певного рангу, тобто співвідношення між рангом слів (R) та величиною тексту, який вони покривають (Π) [Бук 2006]. Результати обрахунків покажуть відсоток покриття тексту найчастотнішими словами. Величину покриття тексту для певного рангу обчислюють як відношення суми абсолютних частот усіх слів з меншими рангами до загальної кількості слів у тексті. У ранговому списку слів за спадом частот бралось послідовно перших за рангом (тобто найчастотніших) декілька (n) позицій, сума (F) їхніх абсолютних частот (F_i) ділилася на загальну кількість слововживань у корпусі, наприклад, ЧС НПКМ ($N = 22948$) і множилася на 100 % (оскільки нам зручно оперувати результатом у відсотках):

$$\Pi = F / N \times 100 \%$$

$$F = F_1 + F_2 + \dots + F_n = \sum_{i=1}^n F_i$$

Наприклад, для перших за рангом 25 слів сума абсолютних частот становить 6733 (визначено за списком слів за спадом

частот), тоді формула набуває вигляду: $\Pi = 6733 / 22948 \times 100\% = 29,3 \%$. Отже, перших за частотністю 25 слів покривають 29,3 % тексту, перших 50 слів – 39,2 %, 100 слів – 50,3 % тексту НПКМ.

За тією ж методикою підраховано співвідношення між рангом слів (*R*) та величиною тексту, який вони покривають (*П*) у ФХ, ПКМ та КМ. Перших за частотністю 25 слів покривають 29,2 % тексту ФХ, 30,5 % тексту ПКМ і 28,5 % КМ. Перших 100 слів – 50 % тексту ФХ, 50,5 % ПКМ і 47,4 % тексту КМ.

Отже, відсоток покриття текстів китайських підручників початкового рівня найчастотнішими словами збільшується зі збільшенням рангу слів.

Найзагальніші статистичні характеристики словникового складу [Бук 2006; 2008] – це обсяг тексту (*N*) та обсяг словника лексем (*V*). Відносними статистичними показниками є: індекс кореляції, індекс словникової різноманітності (*B*), індекс винятковості (*E*), середня повторюваність слова в тексті (*A*). Зазначені статистичні показники є відносними, тому що залежать від обсягу і від типу тексту: при збільшенні обсягу тексту середня повторюваність слів збільшується, а індекс словникової різноманітності зменшується.

Обсяг словника – це величина, яка демонструє, скільки окремих слів зустрілося у відповідних проаналізованих корпусах текстів (*N*).

Табл. 1. Обсяг словника НПКМ, ФХ, ПКМ і КМ

ЧС	Обсяг словника (<i>V</i>)	Обсяг тексту (<i>N</i>)
НПКМ	2764	22323
ФХ	2149	22948
ПКМ	1819	13871
КМ	3767	8675

Отже, найбільший обсяг словника зафіксовано в КМ, найменший – у ПКМ, за ними за зменшенням ідуть НПКМ та ФХ. Можна було б припустити, що чим більший за довжиною проаналізований текст, тим різноманітніший і, відповідно,

більший за обсягом його словник. Але при майже однаковій довжині першого і другого масиву словник першого значно перевищує за обсягом словник другого. Однак третій і четвертий словники все-таки цю думку підтверджують. Тож кількісно лексичний склад підручників з китайської мови початкового рівня істотно відрізняється.

Індекс словникової різноманітності, або лексичної насиченості, – це відношення обсягу словника лексем (V) до обсягу тексту (N), що обчислюється за формулою:

$$B = V / N$$

Проте при вираховуванні цього показника до уваги потрібно брати й обсяг тексту, адже чим більший обсяг, тим більша повторюваність слів і менший показник лексичної насиченості. Якщо довжина тексту дорівнює десять слів, то його лексична насиченість буде в межах 80–100 %. Якщо кількість слів 1 000, то лексична насиченість вже зменшиться до 50 %. Після 200 тисяч слів кількість різних слів вже не збільшується. Тобто чим довший текст, тим меншим буде індекс різноманітності.

Підставивши відповідні дані у формулу, встановлюємо, що в ККМ індекс різноманітності становить 0,17, в НПККМ – 0,12, в ФХ – 0,15, у ПККМ – 0,21.

Результати обчислення індексу різноманітності абсолютно точно корелюють із даними обсягу словника. Найвищий індекс різноманітності в ПККМ, тому що обсяг тексту найменший. Найнижчий індекс різноманітності в НПККМ, тому що обсяг тексту найбільший, середні показники – в ФХ та ККМ.

Середня повторюваність (частота) слова в тексті (A) показує, скільки приблизно разів кожне слово вживається в тексті, тобто це відношення обсягу тексту (N) до обсягу словника лексем (V) – величина, обернена до індексу різноманітності, – обчислюється за формулою:

$$A = N / V$$

Ця величина для ККМ становить 5,92, для НПККМ – 8,30, для ФХ – 6,45, для ПККМ – 4,77. Тобто кожне слово найчастіше зустрічалося в НПККМ (близько 8 раз).

Індекс винятковості, або *harax legomena*, для тексту характеризує варіативність лексики, тобто показує частку тексту, яку займають слова, що трапилися 1 раз, і обчислюється як відношення кількості слів із частотою 1 до обсягу словника. Для тексту обчислюється як відношення кількості *harax legomena* до обсягу тексту ($V1/N$), а для словника – як відношення кількості *harax legomena* до обсягу словника ($V1/V$). Для тексту (E_t), тобто відношення кількості слів із частотою 1 ($V1$) до обсягу тексту (N), обчислюється за формулою:

$$E_t = V1 / N$$

Індекс винятковості для словника (тобто для загальної кількості окремих слів, зведених до початкової форми V) (E_s) обчислюється за формулою:

$$E_s = V1 / V$$

Кількість слів із частотою 1 ($V1$) становить для ККМ 1847, для НПККМ – 1061, ФХ – 875, ПККМ – 891. Підставляємо числа, наприклад, ККМ у формули і множимо на 100 %: $E_t = 1847 / 22323 = 0,0827 \approx 8,27$ %, $E_s = 1847 / 3767 \approx 0,4903 \approx 49$ %.

Аналогічно підраховуємо цей показник для решти словників. Результати обчислень показують, що найбільший індекс винятковості щодо обсягу тексту в підручнику ПККМ, найменший – в НПККМ, щодо обсягу словника найбільший індекс винятковості в ККМ, найменший – в НПККМ. Ці числа – показники варіативності лексики.

Індекс концентрації – величина, протилежна до індексу винятковості, що вказує, яку частку тексту (N) або словника (V) становить високочастотна лексика (з абсолютною частотою 10 і більше). Позначається відповідно $V10_t$ і $V10$. Обчислюється за формулами $V10_t / N$ – індекс концентрації тексту та $V10 / V$ – індекс концентрації словника. Показники цієї величини є найнижчими щодо тексту для ПККМ і, логічно, найвищими для НПККМ. Щодо обсягу словника найбільший індекс концентрації для НПККМ, найменший – у ККМ.

Отже, основні кількісні характеристики лексики НПККМ, ФХ, ППКМ та ККМ унаочнено в таблиці:

Табл. 2. Порівняльна таблиця основних частотних показників НПККМ, ФХ, ППКМ і ККМ

	N	V	V / N	V1 / N	V1 / V	V10т / N	V10 / V
НПККМ	22948	2764	0,120	0,046	0,384	9,59	0,1349
ФХ	13871	2149	0,155	0,063	0,407	0,0016	0,1131
ПККМ	8675	1819	0,210	0,103	0,489	0,0016	0,0113
ККМ	22323	3767	0,169	0,083	0,490	0,0021	0,0876

Тут N – кількість слів у корпусі, V – кількість різних слів (обсяг словника лексем), тобто слів у словнику, V / N – індекс різноманітності, V1 / N – індекс винятковості тексту, V1 / V – індекс винятковості словника, V10т / N – індекс концентрації тексту, V10 / V – індекс концентрації словника.

Отже, за допомогою статистичного аналізу встановлено основні квантитативні характеристики словникового складу НПККМ, ФХ, ППКМ, ККМ, а саме: співвідношення між рангом слова та величиною покриття тексту, кількість різних слів, індекс різноманітності, середню повторюваність слова в тексті, кількість парах *legomena*, індекс винятковості в тексті і в словнику, індекс концентрації. Попри те що слів у корпусі (N) НПККМ удвічі більше, ніж у ППКМ, 22948 і 8675 відповідно, різних слів (V) в НПККМ лише на тисячу більше, ніж у ППКМ, 2764 і 1819. Ці дані вказують на високу частоту повторень нових слів у текстах і діалогах НПККМ, що сприяє їхньому кращому запам'ятовуванню [王凤玲 2020]. Суттєво, що довжина корпусів НПККМ та ККМ фактично однакова, 22948 та 22323, але кількість різних слів у ККМ на 1000 більша, ніж в НПККМ, 3767 і 2764 відповідно, крім того, частка слів, що зустрічаються один раз у ККМ та ППКМ, становить майже половину словника (49 %!!!), водночас в НПККМ та ФХ – лише 38 % та 40 %. Перевага парах *legomena*, тобто слів, що зустрічаються один раз, у підручниках початкового рівня не сприяє засвоєнню активного і пасивного лексичного мінімумів, адже, як відомо, чим слово частіше повторюється в текстах підручника, тим краще запам'ятовується [刘颖 2016]. Що ще цікавіше, кількість лексики, що зустрічається більш ніж 10 разів у ККМ, тобто високочастотна лексика, становить лише

0,98 % словника, водночас в НПККМ – 1,7 %. Промовистими також є кількісні дані індексу винятковості тексту. Статистика, за В. Перебийніс [Перебийніс 2013], будується на так званому “законі великих чисел”, згідно з яким результати статистичних підрахунків тим точніші й вірогідніші, чим більше матеріалу проаналізовано, але збільшення корпусу дає приріст нових слів. Зокрема, у ПККМ кількість слів у корпусі становить 8675, корпус НПККМ удвічі більший, відповідно в НПККМ кількість нових слів більша. Варто також зауважити, що згідно з навчальною програмою “Рівні володіння лексикою та ієрогліфами сучасної китайської мови” [国家汉语... 2001] лексика китайської мови поділяється на чотири ступені: А (甲级词) – 1033 слова, В (乙级词) – 2018 слів, С (丙级词) – 2202 слова, D (丁级词) – 3569 слів. Підручники початкового рівня складаються з лексики ступеня А (甲级词) та В (乙级词), тобто 3015 слів. Однак отримані дані засвідчують, що кількість різних слів (V) у корпусах підручників не відповідає вимогам зазначеної навчальної програми, що є основою при формуванні словникового складу навчальних підручників. Квантитативні характеристики корпусів вказують на незбалансованість лексичного складу підручників з китайської мови початкового рівня в кількісному співвідношенні. Ці характеристики в перспективі корисно буде зіставити з аналогічними квантитативними даними всіх підручників з китайської мови. Отримані кількісні дані можуть бути підтвердженням, що вже така невелика з позиції сучасної корпусної лінгвістики вибірка слів репрезентує основні відмінності в кількісному складі сучасних підручників з китайської мови початкового рівня, а також вказує на відсутність єдиної навчальної системи поглядів щодо формування лексичного складу підручників. Тобто правомірним і перспективним є використання отриманих квантитативних характеристик лексичного складу вже наявних підручників з китайської мови початкового рівня (оскільки цей матеріал перевірений практикою), виданих у КНР, з метою виокремлення і системно-структурного аналізу лексики та укладання частотного словника, що може стати основою при створенні концепції та написанні вітчизняного підручника з китайської мови для українських студентів.

ЛІТЕРАТУРА

- Бук С. Н. **Основи статистичної лінгвістики**. Львів, 2008.
- Бук С. Н. Статистичні характеристики лексики основних функціональних стилів української мови: спроба порівняння // **Лексикографічний бюлетень**. Вип. 13. Київ, 2006.
- Карамнов А. С. Модель создания корпуса учебника английского языка // **Научный диалог**. Вып. 14. Екатеринбург, 2013.
- Перебийніс В. І. **Статистичні методи для лінгвістів**. Вінниця, 2013.
- 白婉莹. “博雅汉语” 副词研究. 北京, 2020.
- 曹漫雯. 新概念英语与新实用汉语课本的对比研究. 北京, 2018.
- 迟惠倩. “博雅汉语” 与 “发展汉语” 初级综合教材比较研究. 北京, 2017.
- 褚琦. 对外汉语初级综合教材 “发展汉语” 与 “新实用汉语课本” 比较研究. 北京, 2018.
- 耿直. “博雅汉语” 初级阶段教材词汇重现率研究. 北京, 2019.
- 国家汉语水平考试委员会办公室考试中心. “汉语水平词汇与汉字等级大纲”. 北京, 2001.
- 郭曙纶, 杨晓惠. 三套对外汉语高级教材课文字词的对比统计与分析. 上海, 2016.
- 侯艳婷. “博雅汉语” 与 “汉语教程” 特殊句型对比研究. 上海, 2018.
- 黄润杉. “汉语教程” 第一册生词英文注释研究. 上海, 2018.
- 江子莹. “发展汉语初级综合” 生字形义关系考察及教学建议. 北京, 2019.
- 李小玲. “发展汉语” 初级阶段系列教材文化因素研究. 北京, 2019.
- 刘诗涵. “博雅汉语” 情态副词调查与分析. 上海, 2016.
- 刘珣. “新实用汉语课本” (I-IV). 北京, 2002.
- 刘颖. “发展汉语” 综合教材新词语收录现状分析. 上海, 2016.
- 任雪梅, 徐晶凝. “博雅汉语初级” (I-II). 北京, 2004.
- 荣继华. “发展汉语初级综合” (I-II). 北京, 2011.
- 孙彦. 对 “新实用汉语课本” (初级部分) 词汇处理情况的考察与分析. 上海, 2019.
- 王凤玲. 从韩国版 “汉语教程” 看汉语教材的本土化改编. 上海, 2020.
- 王淼. “汉语教程” 生词复现情况统计及分析. 北京, 2017.

张宁宁. 新实用汉语课本与博雅汉语（初级）比较研究. 吉林, 2015.

杨寄洲. “汉语教程” (I–III). 北京, 2009.

赵丹慧. “博雅汉语”与“新实用汉语课本”语言点例句对比分析. 上海, 2015.

周晓晔. “博雅汉语”名量词使用情况研究. 北京, 2020.

语料库在线 **Chinese Corpus online**. URL: <http://corpus.zhonghua.yuwen.org/>. (дата звернення: 07.12.2021).

REFERENCES

Buk S. N. (2008), *Osnovy statystychnoyi linhvistyky*, Vydavnychyy tsentr LNU imeni Ivana Franka, L'viv. (In Ukrainian).

Buk S. N. (2006), *Statystychni kharakterystyky leksyky osnovnykh funktsional'nykh styliv ukrayins'koyi movy: sprobа porivnyannya*, in *Leksykohrafichnyy byuletен'*, Vyp. 13, Kyiv, pp. 166–170. (In Ukrainian).

Karamnov A. S. Model' sozdaniya korpusa uchebnika angliyskogo yazyka, in *Nauchnyy dialog*, Vyp. 14, Ekaterinburg, 2013. (In Russian).

Perebyynis V. I. (2013), *Statystychni metody dlya linhvistiv*, NK Publishers, Vinnytsya. (In Ukrainian).

Bai Wanying (2020), “*Boya hanyu*” *fuci yanjiu*, Wenhui chubanshe, Beijing. (In Chinese).

Cao Manwen (2018), *Xin gainian yingyu yu xin shiyong hanyu keben de duibi yanjiu*, Renmin wenxue chubanshe, Beijing. (In Chinese).

Chi Huiqian (2017), “*Boya hanyu*” yu “*Fazhan hanyu*” *chuji zonghe jiaocai cankao yanjiu*, Beijing chubanshe, Beijing. (In Chinese).

Chu Qi (2018), *Zhongxue hanyu chuji zonghe jiaocai “Fazhan hanyu” yu “Xin shiyong hanyu keben” bijiao yanjiu*, Zhongguo shehui kexue chubanshe, Beijing. (In Chinese).

Geng Zhi (2019), “*Boya hanyu*” *chuji jieduan jiaocai fuxi*, Beijing chubanshe, Beijing. (In Chinese).

Guojia hanyu shuiping kaoshi weiyuanhui bangongshi kaoshi zhongxin (2001), *Hanyu shuiping cihui yu hanzi dengji dagang*, Beijing jingji kexue chubanshe, Beijing. (In Chinese).

Guo Shulun and Yang Xiaohui (2016), *San tao duiwai hanyu jiaocai kewen zici de duibi tongjiyu fenxi*, Shanghai chubanshe, Shanghai. (In Chinese).

Hou Yanting (2018), “Boya hanyu” yu “Hanyu jiaocheng” teshu ju xing duibi yanjiu, Shanghai, Shanghai chubanshe, Shanghai. (In Chinese).

Huang Runshan (2018), “Hanyu jiaocheng” diyi ce shengci zhongwen zhushiyanjia, Zhongguo shehui kexue chubanshe, Shanghai. (In Chinese).

Jiang Ziyang (2019), “Fazhan hanyu chuji zonghe” sheng zixing yiguanxiluyou ji jiaoxue jianyi, Renmin wenxue chubanshe, Beijing. (In Chinese).

Li Xiaoling (2019), “Fazhan hanyu” chuji xilie jiaocai wenhua yinsu yanjiu, Wenhui chubanshe, Beijing. (In Chinese).

Liu Shihan (2016), “Boya hanyu” qingtai fuci diaocha yu fenxi, Shanghai chubanshe, Shanghai. (In Chinese).

Liu Xun (2002), “Xin shiyong hanyu keben” (I–IV), Beijing yuyan daxue chubanshe, Beijing. (In Chinese).

Liu Ying (2016), “Hanyu fazhan” zonghe jiaocai xin ci shoulu xianzhuang fenxi, Shanghai chubanshe, Shanghai. (In Chinese).

Ren Xuemei and Xu Jingning (2004), “Boya hanyu chuji” (I–II), Beijing daxue chubanshe, Beijing. (In Chinese).

Rong Jihua (2011), “Fazhan hanyu chuji zonghe” (I–II), Beijing yuyan daxue chubanshe, Beijing. (In Chinese).

Sun Yan (2019), *Dui “Xin shiyong hanyu keben” (chuji bufen) cihuichuli qingkuang de zhuti yu fenxi*, Shanghai chubanshe, Shanghai. (In Chinese).

Wang Fengling (2020), *Cong hanguo ban “hanyu jiaocheng” kan hanyu jiaocai de bentu hua gaizao*, Shanghai chubanshe, Shanghai. (In Chinese).

Wang Miao (2017), “Hanyu jiaocheng” shengci fu xian qingkuang tongjiji fenxi, Zhongguo shehui kexue chubanshe, Beijing. (In Chinese).

Yang Jizhou (2009), “Hanyu jiaocheng” (I–III), Beijing yuyan daxue chubanshe, Beijing. (In Chinese).

Zhang Ning (2015), *Xin shiyong hanyu keben yu boya hanyu (chuji) bijiao yanjiu*, Jilin chubanshe, Jilin. (In Chinese).

Zhao Danhui (2015), “Boya hanyu” yu “Xin shiyong hanyu keben” yuyan dian fanyiduibi fenxi, Shanghai yanjiu chubanshe, Shanghai. (In Chinese).

Zhou Xiaoye (2020), “Boya hanyu” ming liangci shiyong qingkuang yanjiu, Zhongyang bianyi chubanshe, Beijing. (In Chinese).

Chinese Corpus online (2004), available at: <http://corpus.zhonghuayuwen.org/> (accessed 7 December 2021).

Н. С. Ісаєва, І. В. Ситник

Квантитативні характеристики лексичного складу підручників з китайської мови початкового рівня

Безперечно є важливість застосування квантитативних методів у мовознавстві, адже значну кількість мовних закономірностей треба трактувати не як абсолютні закони, а як правила, які впливають зі статистичної ймовірності. Будова мови та її функціонування в мовленні підлягають дії статистичних законів, тому досліджувати мову необхідно за допомогою чітких статистичних методів для аналізу мовного матеріалу та його систематизації. Застосування статистичних методів дає змогу здобувати точні кількісні дані і на їхній підставі робити лінгвістичні висновки та простежувати певні тенденції у функціонуванні мовних одиниць у мовленні. У статті представлено результати статистичного дослідження лексичного складу найпопулярніших підручників з китайської мови початкового рівня, а саме: “Новий практичний курс китайської мови” “新实用汉语课本”, “Комплексний курс китайської мови *Фачжсань ханьюй*: початковий рівень” “发展汉语初级综合”, “Поглиблений курс китайської мови. Початковий рівень” “博雅汉语初级”, “Курс китайської мови” (початковий рівень) “汉语教程”. Встановлено, що в сучасній китайській лінгвістиці та лінгводидактиці особливо актуальними є проблема кількісного і якісного відбору навчального матеріалу та проблема відсутності практичних об’єктивних даних про одиниці мови, які використовуються в підручниках з китайської мови різних рівнів. Зазначено, що частотний словник як модель організації лексики допомагає вирішити проблему невизначеності словникового запасу слів. У статті зафіксовано специфіку квантитативних досліджень вищезазначених підручників у китайській лінгвістиці. Систематизовано та описано кожний етап укладання власного частотного словника підручників з китайської мови початкового рівня. Представлено результати застосування методу статистичного аналізу. Встановлено основні квантитативні характеристики словникового складу вищезазначених підручників: обсяг словника лексем, індекс словникової різноманітності,

індекс винятковості, середня повторюваність слова в тексті тощо. Доведено, що студенти, використовуючи різні підручники початкового рівня, вивчають абсолютно різну кількість слів, що унеможлиблює справедливе оцінювання знань студентів початкового рівня та труднощів у процесі переходу на середній рівень володіння китайською мовою. Зосереджено увагу на тому, що існує потреба в дослідженні словникового складу підручників з китайської мови початкового рівня з метою виокремлення та системно-структурного аналізу спільної лексики та укладання частотного, що може стати основою для написання майбутнього підручника з китайської мови для українських студентів.

Ключові слова: китайська мова, статистичний аналіз, квантитативні характеристики, підручник, початковий рівень

Стаття надійшла до редакції 16.12.2021