

Лінгвостатистичний аналіз лексики китайськомовного педагогічного дискурсу

I. В. Ситник

Інститут філології, Київський національний університет імені Тараса Шевченка, м. Київ, Україна
Corresponding author. E-mail: sytnykilona@gmail.com

Paper received 19.11.21; Accepted for publication 29.11.21.

<https://doi.org/10.31174/SEND-Ph2021-260IX76-13>

Анотація. В статті розглянуто специфіку використання методу квантитативного аналізу в мовознавстві на прикладі китайської мови. Проаналізовано основні типологічні особливості китайської мови, які необхідно враховувати при лінгвостатистичному аналізі лексики китайськомовного педагогічного дискурсу для забезпечення достовірності отриманих статистичних даних. Відзначено, що результатом використання лінгвостатистичного методу в дослідженні мовних одиниць китайськомовного педагогічного дискурсу є частотний словник.

Ключові слова: китайська мова, типологічні особливості, лексика, лінгвостатистичний аналіз, частотний словник.

Вступ. Китайська мова є однією з найдавніших мов світу, якою спілкується наразі більше мільярда людей. В Україні з кожним роком збільшується кількість охочих вивчати китайську мову. Про зростання популярності китайської мови в Україні свідчить щорічне збільшення кількості абітурієнтів, які бажають навчатися на китаєзнавчих відділеннях університетів. Однак, на сучасному етапі викладання китайської мови педагоги часто зіштовхуються з проблемою відсутності належної лінгвістичної теоретичної бази, необхідної для вирішення питання кількісного та якісного відбору навчального матеріалу, а також з відсутністю практичних об'єктивних даних про одиниці мови, які можуть бути використані на різних рівнях вивчення китайської мови. Вирішити вищезазначені проблеми можна за допомогою лінгвостатистичного методу дослідження.

Короткий огляд публікацій по темі. Лінгвостатистичні дослідження лексики китайськомовного педагогічного дискурсу можна розділити на дві групи: перша група науковців (Сунь Янь (孙彦), Чу Ці (褚琦), Цзян Цзіїн (江子莹), Лі Сяолін (李小玲), Лю Ін (刘颖), Фен Цай (丰彩), Тань Сін (谭欣)) комплексно досліджує два різних комплекти навчальних матеріалів, порівнює різні аспекти, такі як структура, граматичні конструкції, теми текстів та словниковий склад тощо; друга група (Лю Хишань (刘诗涵), Чжоу Сяо (周晓晔), Бай Ваньїн (白婉莹), Хуан Цзін (黄婧), Пен Мінь (彭敏), Ян Даньдань (杨丹丹), Ген Чжи (耿直)) здійснює порівняльний аналіз двох різних підручників у межах одного аспекту, наприклад, словникового складу. Це перша спроба у вітчизняній науці за допомогою лінгвостатистичного методу здійснити аналіз китайськомовного педагогічного дискурсу, тож дана розвідка може посприяти глибшому розумінню особливостей китайської мови, які необхідно врахувати при лінгвостатистичному аналізі.

Мета наукової розвідки – проаналізувати особливості китайської мови, для здійснення лінгвостатистичного аналізу з метою отримання достовірних статистичних даних.

Матеріали та методи. Дослідження включає описовий, філологічний, лінгвостатистичний методи, а також аналіз, класифікація та узагальнення. Матеріалом є лексика китайськомовного педагогічного дискурсу.

Результати та їх обговорення. Лінгвостатистичний

метод – це сукупність статистичних методик, прийомів і процедур, які використовуються для отримання кількісних даних про мовні явища з метою виявлення статистичних закономірностей функціонування одиниць мови та мовлення [6, с. 87]. Застосування методів статистики у дослідженні лінгвістичних об'єктів є одним з актуальних напрямів сучасного мовознавства, оскільки дає підстави глибше зрозуміти багато явищ мови та мовлення. Квантитативні методи дозволяють отримати кількісну інтерпретацію лінгвістики тексту, яка доповнює якісний лінгвістичний аналіз. Якісний аналіз мови – це категоризація мови, тобто виділення у ній окремих класів явищ, об'єднаних якісними ознаками [4, с. 22]. Однак, будь-яка категоризація пов'язана з квантифікацією мови, тобто її кількісним аналізом.

Слідом за Р. Мельничук, вважаємо, що тенденція щодо застосування статистичних прийомів при дослідженні лінгвістичних явищ обґрунтовується такими чинниками: «об'єктивною властивістю мови і кількісних характеристик; внутрішньою залежністю, яка існує між якісними і кількісними характеристиками мовної структури; підвладністю частоти різних елементів мови у мовленнєвому потоці тим чи іншим статистичним законам» [5, с. 36]. Результатом використання лінгвостатистичного методу в дослідженні мовних одиниць є частотний словник (далі ЧС). Такі словники дають уявлення про статистичну структуру тексту, на основі якого словник укладений, адже ЧС є статистичною моделлю тексту.

Однак, при лінгвостатистичному аналізі китайської мови необхідно враховувати її типологічні особливості. Китайська мова належить до китайської групи сино-тибетської мовної сім'ї, за морфологічною класифікацією належить до кореневих (ізолюючих) мов. Здійснений з урахуванням основних особливостей китайської мови лінгвостатистичний аналіз китайської мови може стати джерелом достовірних кількісних даних.

Диференціації термінів «ієрогліф», «морфема», «склад», «слово» в китайській мові.

Центральне місце в лексичній системі китайської мови займає слово. Однак, межі китайського слова є розмитими та неоднозначними тому, що часто отожднюється із складом, морфемою або ієрогліфом. В китайській мові, як і в українській мові, слово поділяється на склади. Однак, в українській мові складу властивий лише план вираження, це звукова одиниця з допомо-

гою якої будують слова. Водночас в китайській мові, як правило, склад має і план вираження, і план значення. Тобто, в китайській мові поділ слова на склади співпадає із морфологічним поділом, тому морфему називають складоморфемою. Загальноприйнятим є розуміння морфеми як складової частини слова, що не може виступати самостійною одиницею комунікації. Це трактування морфеми не може застосовуватись в межах китайської мови. Адже особливістю китайською мови є те, що майже всі морфеми водночас є односкладовими словами, тому важко провести межу між морфемою та словом. На письмі морфема фіксується окремим знаком – ієрогліфом “字”. А. Семенас визначає “字” «ієрогліф» як первинну лексему [7, с. 51]. Первинним лексемам властива граматична визначеність, їх вимовляють певним тоном і вони є мовною реальністю, що пов'язано з особливим типом китайської мови і мовного мислення китайців. Більшість первинних лексем передається одним ієрогліфом і одним складом, а багатоскладові лексичні одиниці вважаються складними за своєю морфологічною структурою, тобто вони складаються з декількох первинних лексем. Такі складні слова і стійкі словосполучення, в свою чергу, утворюють лексичні комплекси за певними словотворчим типам або словотворчим моделям. За такими моделями будуються як загальноживані слова китайської мови, так і вузькоспеціалізовані терміни і термінологічні сполучення. Якщо в українській мові первинним є слово, а вторинним морфема, то в китайській мові первинним є ієрогліфів (первинна лексема), а вторинним слово [7, с. 52].

В китайській лінгвістиці є два підходи до визначення терміна «слово». В межах першого підходу, на думку китайських дослідників (Чжан Шичжао (章士钊), Чжао Юаньжень (赵元任), Хуан Бочжун (黄波忠), Ляо Сюйдун (廖徐东) та інші) в китайській мові відсутнє поняття «слово» у загальноприйнятому розумінні, замість нього наявний термін “字” «ієрогліф». Ієрогліф являє собою мінімальну одиницю з двоїстою природою лінгвістичного знаку – план вираження і план значення [3, с. 171].

Хуан Бочжун (黄波忠) та Ляо Сюйдун (廖徐东) визначають ієрогліфи як систему знаків, за допомогою якої записується мова [11, с. 55]. Чжан Шичжао (章士钊) поєднує терміни «ієрогліф» та «слово»: «Один ієрогліф може бути словом, але слово не обов'язково складається з одного ієрогліфа» [12, с. 183]. Чжао Юаньжень (赵元任) зазначає, що «<...> термін «слово» в мовах індоєвропейської сім'ї не має відповідника в китайській мові <...>, слово – це термін європейської лінгвістики, що не тотожне ієрогліфу» [13, с. 42].

Представники другого підходу (Ху Юйшу (胡裕树), Гао Мінкай (高名凯), Фу Хуайцін (符淮青)) не розглядають “字” «ієрогліф» як основну одиницю мови, а послуговуються терміном “词” «слово». Гао Мінкай (高名凯) визначає “词” «слово» як основну одиницю лексики або будівельний матеріал мови [14, с. 48]. Фу Хуайцін (符淮青) розуміє “词” «слово» як мінімальну одиницю, яка може окремо використовуватися або утворювати речення [15, с. 162].

У традиційній граматиці китайської мови визначено, що “词” «слово» являє собою певний сенс, має фіксовану фонетичну форму, воно є мінімальною одиницею, яка може вживатися самостійно» [16, с. 37].

Явище омонімії в китайській мові.

Ще однією характерною типологічною особливістю китайської мови є відсутність однозначних зв'язків між графічною та фонетичною репрезентацією одного і того самого ієрогліфа. Саме особлива фонетична організація китайської мови є основною причиною омонімії. Китайська мова – одна з мов із найбільшою кількістю омонімів. Омоніми (同音词) – слова однієї частини мови або їхні окремі граматичні форми, а також стійкі словосполучення, морфеми, синтаксичні конструкції, що за однакового звучання (або написання) мають абсолютно різні значення [7, с. 89]. При цьому під однаковим звучанням слід розуміти не тільки ідентичний склад фонем, але також один і той же тон. Тон – це маніпуляція з висотою голосного звука, яка здійснюється за допомогою зміни швидкості коливань голосових зв'язок [1, с. 12]. Тони виконують смислорозрізнявальну функцію. У китайській мові склад завжди вимовляється відповідним тоном, а неправильна вимова тонів спричиняє зміну значення слів [1, с. 78].

У китайській мові омонімія властива як односкладовим, так і двоскладовим і навіть трискладовим словам. Більшість омонімів китайської мови графічно диференційовані. Одній односкладовій (а також багатоскладовій) лексичній одиниці, виділеній на фонетичному рівні, зазвичай, відповідають на графічному рівні декілька різних ієрогліфів. Для того, щоб зрозуміти особливості китайських омонімів потрібно розглянути їх класифікацію. А. Хаматова виділяє три типи омонімів [8, с. 45]:

1) Слова збігаються за звучанням і написанням (同音同形词). У цьому випадку один і той же ієрогліф записується та звучить однаково. Наприклад, 词 сі «слово», та 词 сі «назва стародавнього поетичного жанру ци».

2) Слова збігаються за звучанням, але не збігаються з написанням (同音异形词). Це фонетичні омоніми, або омофони. Наприклад, 班 bān «група» і 斑 bān «пляма», 八 ba «вісім» і 扒 ba «викопувати». У китайській мові найбільш поширений саме цей тип омонімії.

3) Слова мають однакове написання, але різне звучання (胃形异音词). Це омографи. Наприклад, 还 huán «повертати», 重 zhòng «важкий» і 重 chóng «знову», 长 cháng «довгий» і 长 zhǎng «рости».

У нашому дослідженні при укладанні ЧС ми здійснювали усунення омонімії шляхом додавання до одного з омонімічної пари умовної позначки при розмітці і сегментації текстів у корпусі за допомогою мережевої програми Chinese Corpus online – Automatic POS-tagging and segmentation (分词和词性标注) [17]. Таким чином ці слова графічно розрізнялися, і програма рахувала частоту вживання кожного з них окремо. Наприклад, в укладеному нами корпусі є чотири 下: дієслово 下 (下课 завершувати урок, 下班 завершувати роботу), прийменник 下 (下 під), іменник (下面 низ), рахівне слово (看一下 подивитись).

Кількісне співвідношення між складами в китайській мові.

Слова в китайській мові можуть бути простими або складеними. Прості слова складаються з однієї складоморфеми, тобто є односкладовими (наприклад, 天 «небо», 去 «йти» 他 «він») або багатоскладовими (往往 «часто», 咖啡 «кава»). Складені слова складаються з декількох морфем. У китайській мові наявні різні типи складених слів, що класифікуються за їх внутрішньою структурою, а саме: атрибутивний (偏正型), копулятивний (联合型), дієслівно-об'єктний (述宾型), дієслівно-результативний (补充型), суб'єкт-предикативний (主谓型), афіксація (附加式), редуплікація (重叠式), аббревіація (简称), цифрові скорочення (数词缩语), запозичення (外来词) [2, с. 16].

В китайській мові іноді неможливо провести межу між морфемами та словами, з одного боку, та між словами і фразами, з іншого боку. Зараз існує цілий комплекс критеріїв для визначення слів різних типів, наприклад, слово «фонологічне», «морфологічне», «лексичне», «синтаксичне», «граматичне», «семантичне», «соціологічне», «психологічне», «фонологічне», «просодичне». Так, наприклад, на основі аналізу таких критеріїв Роберт Діксон та Олександра Айхенвальд [9, с. 15] пропонують чітко розділяти фонологічне та граматичне слово, які можуть збігатися або не збігатися. Хоча автори визнають, що критерії, які вони застосовують для визначення фонологічного слова, застосовуються не в кожній мові, вони пропонують три «універсальні критерії», що визначають граматичне слово: «Грамматичне слово складається з ряду граматичних елементів, які (а) завжди використовуються разом, а не розкидані (критерій згуртованості); (б) використовуються у фіксованому порядку; (с) мають узгодженість і значення» [9, с. 13].

Слова, що мають дієслівно-об'єктну (述宾型) внутрішню структуру не відповідають жодному з цих критеріїв. Морфеми, що утворюють такі слова, можуть не тільки бути розкидані по реченню, але й не використовуватись разом (наприклад, 睡觉 «спати»: 睡了一天觉 «спав цілий день»), вони можуть використовуватись навіть у зворотному порядку (наприклад, 上学 «ходити до школи»: 学, 我爱上就上, 不爱上就不上, 谁管得着. «Школа, якщо мені подобається, то я піду, якщо не подобається, то не піду, хто має право втручатися»). Р. Діксон та О. Айхенвальд [9, с. 16] зазначають, що критерій (с) означає, що, оскільки слово має свою узгодженість і значення, носії мови «можуть говорити про слово (але навряд чи говорять про морфему)». Цей критерій не застосовується в китайській мові не лише тому, що морфеми в китайській можуть бути узгодженими та мати значення (односкладове слово складається з однієї морфеми), а й тому, що навіть зв'язані морфеми в розділених словах, таких як 请客 «запросити гостя (гостей)» про них можна «говорити», наприклад, 请了一桌客 «запросив стіл гостей». За словами Р. Діксона та О. Айхенвальд [9, с. 16], «(граматичне) слово поєднує морфологію та синтаксис. Морфологія

має справу зі складом слів, тоді як синтаксис має справу з поєднанням слів». Однак у китайській мові внутрішні структури слів подібні до синтаксичних структур, як це можна побачити у розглянутих вище типах китайських складних слів.

Послугуючись результатами досліджень американського дослідника Сун Цзяна [10, с. 63], при укладанні ЧС ми не здійснювали розділення деяких складених іменників шляхом додавання умовної позначки при розмітці і сегментації текстів у корпусі за допомогою мережевої програми Chinese Corpus online – Automatic POS-tagging and segmentation (分词和词性标注) [17]. Таким чином ці слова графічно не розрізнялися, і програма рахувала частоту вживання кожного з них як одного слова. Наприклад, слова 卖 «продавати» та 完 «завершувати» Сун Цзян у своєму дослідженні розрізняє як окремі слова у складеному іменнику 卖完了 «розпроданий». У нашій роботі, послугуючись мережевою програмою Chinese Corpus online – Frequency statistics of words (字词频率统计) [17], складені іменники рахувались як одне слово 卖完.

Визначення частиномовної приналежності слів в китайській мові.

У китайській мові слова не мають граматичної афіксації, словозмін та морфологічно закріплених частиномовних характеристик, що дозволяє лексичній одиниці без жодних змін виступати в різних семантико-синтаксичних умовах у ролі різних лексико-граматичних класів. Тобто відбувається перехід слова з однієї частини мови в іншу або його використання у функції іншої частини мови. Функційна транспозиція – це підстава одночасного поєднання в одній одиниці функції декількох граматичних класів. Китайські вчені формують цю проблему у стилій формі «一次多类» (одне слово – багато категорій).

Висновки. Досліджено, що при лінгвостатистичному аналізі лексики китайськомовного педагогічного дискурсу для забезпечення достовірності отриманих статистичних даних необхідно враховувати типологічні особливості китайської мови, а саме: 1) відсутність «слова» у загальноприйнятому значенні цього терміну, і, як наслідок, складність виділення одиниць лексичної системи; 2) кількісний склад лексичних одиниць китайської мови, що пов'язано з відсутністю узгодженості в питанні визначення довжини китайського «слова»; 3) дворівнева структура китайської лексики – відсутність однозначної відповідності між графічною на фонетичною формами репрезентації мовних одиниць; 4) проблема відсутності формально виражених частиномовних характеристик; 5) відсутність форм словозміни, тому слово не має в собі жодних показників синтаксичного зв'язку з іншими словами у реченні; 6) синтаксичний зв'язок між словами будуються на основі незмінного порядку слів у реченні; 7) граматичні значення виражаються способом прилягання одних слів до інших; 8) тонізація як ключове явище смислотворення та смислорозрізнення.

ЛІТЕРАТУРА

1. Кірносова Н. А. Практична фонетика китайської мови. Київ : НАККИМ, 2010. 96 с.
2. Козоріз О. П. Словосполучення та словотвір у сучасній китайській мові. Вісник Київського національного університету імені Тараса Шевченка, 2011. С. 16–18.
3. Курдюмов В. А. Курс китайского языка. Теоретическая грамматика. Москва : Цитадель-Трейд, 2005. 576 с.
4. Левицкий В.В. Квантитативные методы в лингвистике. Винница : Нова книга, 2007. 264 с.
5. Мельничук Р. І. Методи у дослідженні фоносемантичних явищ. Science and education a new dimension. Budapest. Philology, IV(22), Issue: 99, 2016. P. 36–39.
6. Селіванова О.О. Сучасна лінгвістика: термінологічна енциклопедія. Полтава : Довкілля-К, 2008. 844 с.
7. Семенов А.Л. Лексикология современного китайского языка. Москва : Наука, 1992. 314 с.
8. Хаматова А. А. Омонимия в современном китайском языке: Учеб. пособие. Владивосток : ДВГУ, 1981. 127 с.
9. R. M. W. Dixon, Alexandra Y. Aikhenvald. Word: a cross-linguistic typology. Cambridge : Cambridge University Press, 2002. 290 p.
10. Song Jiang. Defining the So-Called 'Core Vocabulary'. A Case Study of Chinese Textbooks. Journal of Chinese Language and Computing, Issue: 16(1), 2004. P. 63–71.
11. 黄波忠, 廖徐东. 简要介绍一下当代中文问题的答案. 外语教学, 1996. № 4. 页 55–70.
12. 章士钊. 汉字的概念、结构和节奏. 江苏社会科学, 2010. № 2. 页 183–188.
13. 赵元任. 语言学模型. 外语与外语教学, 1994. 165 页
14. 高名凯. 汉语语法论. 科学出版社, 1957. 544 页
15. 符淮青. 现代汉语词汇. 北京大学出版社, 2004. 254 页
16. 胡裕树. 现代汉语. 上海教育出版社, 1995. 564 页
17. 语料库在线 Chinese Corpus online [Elektronний ресурс] – Режим доступу до ресурсу: <http://corpus.zhonghuayuwen.org/>.

REFERENCES

1. Kirnosova N. Practical phonetics of the Chinese language. Kyiv: NAKKIM, 2010. 96 p.
2. Kozoriz O. Phrases and word formation in modern Chinese. Bulletin of Taras Shevchenko National University of Kyiv, 2011. P. 16–18.
3. Kurdyumov V. Chinese language course. Theoretical grammar. Moscow: Citadel-Trade, 2005. 576 p.
4. Levitsky V. Quantitative methods in linguistics. Vinnitsa: Nova knjiga, 2007. 264 p.
5. Melnychuk R. I. Methods in phonosemantic research. Science and education a new dimension. Budapest. Philology, IV(22), Issue: 99, 2016. P. 36–39.
6. Selivanova, O.O. Linguistic encyclopedia. Poltava, 2010. 844 p.
7. Semenas A. Lexicology of modern Chinese. Moscow: Nauka, 1992. 314 p.
8. Khamatova A. Homonymy in modern Chinese: Textbook. Vladivostok: DVGU, 1981. 127 p.
11. Huang Bozhong, Liao Xudong. Brief introduce of the answers to contemporary Chinese questions. Foreign language teaching, 1996. № 4. P. 55–70.
12. Zhang Shizhao. The concept, structure and rhythm of Chinese characters. Jiangsu Social Sciences, 2010. № 2. P. 183–188.
13. Zhao Yuanren. Linguistic Models. Foreign Languages and Foreign Language Teaching, 1994. 165 p.
14. Gao Mingkai. Chinese Grammar. Science Press, 1957. 544 p.
15. Fu Huaiqing. Modern Chinese Vocabulary. Peking University Press, 2004. 254 p.
16. Hu Yushu. Modern Chinese. Shanghai Education Press, 1995. 564 p.
17. Chinese Corpus online [E. source]. Access mode: <http://corpus.zhonghuayuwen.org/>.

Lingvostatistical analysis of Chinese pedagogical discourse

I. Sytnyk

Abstract. The present paper aims to define the specifics of quantitative analysis usage in linguistics on the example of the Chinese language. The main typological features of the Chinese language are presented and analyzed. It has been pointed out that Chinese language typological features must be taken into account in lingvostatistical analysis of Chinese pedagogical discourse lexis to ensure the reliability of the obtained statistical data. It is fundamental to note that the result of lingvostatistical analysis is a frequency dictionary.

Keywords: Chinese language, typological features, lexis, lingvostatistical analysis, frequency dictionary.