УДК 81'373.1

УКЛАДАННЯ ЧАСТОТНОГО СЛОВНИКА КОРЕЙСЬКОЇ МОВИ

Білошицький Костянтин Миколайович,
студ.
Київський національний університет імені Тараса Шевченка

У статті розглядаються етапи укладання частотного морфемника корейської мови. Проведено аналіз корпусу, на матеріалі якого було укладено частотний морфемник. Описано його роль і місце у сучасній лінгводидактиці.

Ключові слова: *частотний словник, корпус, корейська мова.*

1) Частотні словники існують для багатьох мов світу. Їх ефективно використовують на практиці: у викладанні, при створені словників-мінімумів, підручників, для розв'язання інших лінгвістичних задач. У навчанні мов ефективність такої методики неодноразово доведена [4].

При укладанні нових підручників з іноземних мов щораз звичайнішим стає вибір словника на підставі частотних списків слів. Ця методика значно прискорює процес вивчення мови порівняно з іншими. Укладання підручників з використанням частотності лексики, як основного критерію, успішно застосовують в Угорщині, Польщі, Італії, Франції. Така методика дозволяє учням читати оригінальні тексти вже на першому семестрі навчання, що саме по собі свідчить про її високу ефективність.

Було проведено дослідження на основі частотних словників таких мов, як англійська, німецька, які показали, що читачеві, який хоче прочитати і перекласти іноземний текст зі спеціальності, яка його цікавить, зовсім не обов'язково володіти усім багатством мови, якою написаний текст. Достатньо активно володіти певним запасом службових, загальновживаних слів і граматичних морфем, а також пасивно знати певну кількість менш вживаних ключових слів і зворотів. Причому досягти перші дві мети якраз дає змогу частотний словник.

2) Для корейської мови дотепер не існувало частотного словника. А він потрібен тому, що є незамінним матеріалом при вивчені іноземної мови, особливо на початковому етапі. Сучасною наукою встановлено, що запас у 1000 найуживаніших слів, скажімо, для російської мови, забезпечує розуміння пересічного тексту на 76 % [1].

Першим етапом укладання частотного списку є відбір достатньо різноманітних уривків тексту, які формують корпус майбутнього словника. Корпус для укладання частотного словника корейської мови було сформовано за такими параметрами. Обсяг вибірки становить 1 056 938 слововживань. Із 1500 відібраних джерел, було взято уривки текстів, жоден з яких за обсягом не перевищує 800 слововживань. Вибірка проводилася з використанням виключно Інтернет джерел [8; 9; 10; 11; 12; 13; 14; 15; 16; 17].

Об'єм корпусу відмінний від словника до словника. Як відомо, найперший статистичний словник мови був задуманий і одноосібно здійснений у Німеччині Ф. Кедінгом ще у 1898 році на корпусі з 11 млн. слів, і включив тексти з 290 джерел від художньої літератури до стенограм дебатів у парламенті [2].

Пізніші словники створені на значно менших за обсягом корпусах. Так, класичні частотні словники англійської мови Е. Трондайка [8], російської мови К. Йосельсона [6] і Л.Н. Засоріної [1] побудовані на корпусі в 1 млн. словоформ. Російський

словник Е. Штейнфельдт [5] і фінський П. Сауконена [7] – на корпусі в 400 тис. словоформ.

Принципи формування корпусу також відмінні від словника до словника. Основною вимогою залишається жанрова різноманітність корпусу. До корпусу прийнято включати тексти з художньої прози, публіцистичні, комунікативні, наукові, газетно-інформаційні. Проте конкретна частка кожного жанру досить відмінна в залежності від словника.

Співвідношення жанрів письмових текстів у корпусах різних частотних словників (рис. 1):

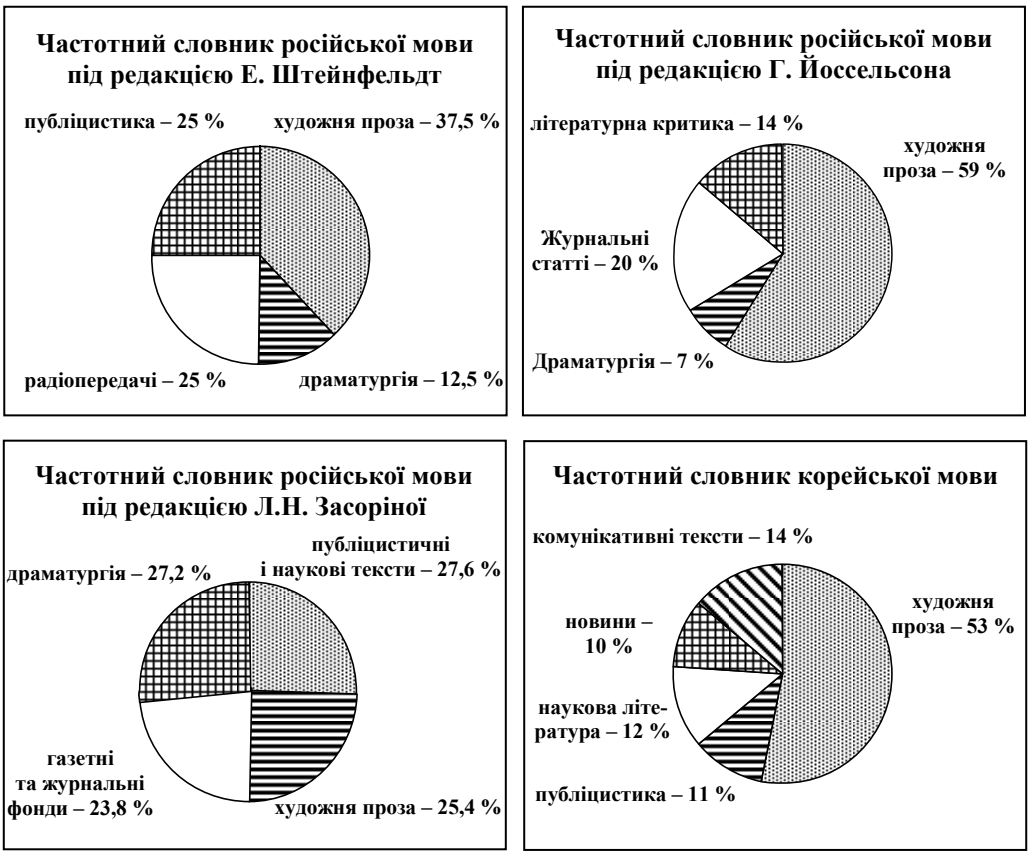

Рис. 1

До корпусу частотного словника корейської мови було включено уривки текстів з художньої літератури – 53 %. Іншу половину майже порівну поділили між собою публіцистика – 11 %, наукова література – 13 %, новини – 10 % і комунікативні тексти – 14 %. Основне ядро комунікативних текстів склали дебати, різного типу обговорення і просто діалоги та інтерв'ю.

Часовий діапазон текстів корпусу обмежено матеріалами 1990-2008 років.

Таким чином, було створено презентабельний корпус корейської мови, який відображає її сучасну письмову норму.

3) Укладання частотних списків слів корейської мови на основі сформованого корпусу проходило в два етапи. На першому етапі отримано частотний словник словоформ, на другому – частотний реєстр морфем, з яких вони складаються.

Для створення списку словоформ було використано авторську програму Б.А. Рудого *Text analyzer* [3]. За допомогою цієї комп'ютерної програми було опрацьовано мільйонний масив тексту, і отримано реєстр словоформ, упорядкованих за спадною частотністю.

Половину всіх слововживань у корпусі сумарно покриває більше ніж 2 тис. одиниць нерафінованого списку. З цілком зрозумілих причин цей показник, порівняно з частотними словниками лексем є низьким. Для лексем він коливається в межах 230 слів. Цей список є лише проміжним етапом дослідження. Словоформи не можуть мати прямої цінності для навчання, бо віддзеркалюють відразу дві характеристики, вжитих у текстів мовних засобів – частотний словник і частотну граматику.

Другим етапом укладання словника стало створення списку морфем. Щоб відрафінувати список необхідно залишити основи і видалити формативи. Для розв'язання цієї задачі, на підставі нормативної граматики, було укладено список граматичних морфем корейської мови. Їх виявилося 1095. Вони були занесені до програми Text analyzer і послідовно відділені від основ, після чого тотожні основи зведені в лексему з вказівкою частоти вживання. Таким чином було отримано два нові списки: частотний список основ слів і частотний список граматичних формативів, реально вжитих у корпусі. При цьому з'ясувалося, що зі списку у 1095 граматичних морфем 356 не зустрілося жодного разу у нашому корпусі з мільйону слів. Більша половина з усіх граматичних морфем не подолали бар'єру у 20 вживань. У такий спосіб вперше було отримано статистичні вірогідні дані для корейської мови про частотність граматичних морфем, що має першорядне значення для обґрунтування відбору граматичних тем при укладанні підручників корейської мови.

У підсумку було отримано три частотні списки: реєстр словоформ, і створені на його основі реєстр основ слів і реєстр граматичних морфем.

Ці списки були відредаговані вручну. Кожну основу було ідентифіковано, зведено в один список всі випадки вживання у списку словоформ приплюсовано до аналогічних основ і розміщено у спадному порядку у реєстрі. Після цієї оброки список набуває завершальної форми.

У процесі редагування списку серед першої тисячі основ було виявлено декілька ієрогліфів. Як відомо, китайська ієрогліфіка використовується в корейських газетах. Слова, зображені ієрогліфами, вважаються порівняно часто вживаними. Такі випадки ієрогліфічного написання замінені написанням корейським алфавітом (хангилем) і зберегли свої місця у реєстрі. Кількість вживань даного слова, записаного ієрогліфом плюсувалася до кількості вживань того ж слова у стандартному записі хангилем, тим самим підвищуючи його позиції у частотному списку.

На рис. 2 показано фрагмент відредагованого рафінованого списку. Для порівняння фрагменти частотних – словників інших мов: української, англійської, російської.

R	Лексема	R	Word	R	Лексема	R	Морфема	Переклад
1	**в (у)**	1	**the**	1	и	1	하	**Роб***
2	**і (й)**	2	**be**	2	в	2	에	**В**
3	**на**	3	**of**	3	не	3	이	**Це / бу****
4	**з (із, зі)**	4	**and**	4	он	4	고	**І**
5	**не**	5	**a**	5	на	5	로	**В (до)**
6	**що**	6	**in**	6	я	6	있	**Бу****
7	**бути**	7	**to**	7	что	7	도	**І**
8	**до**	8	**have**	8	тот	8	과/와	**І**
9	**який**	9	**it**	9	быть	9	되	**Ста*****
10	**та**	10	**to**	10	с	10	것	**Річ**
11	**за**	11	**for**	11	а	11	그	**Це**
12	**той**	12	**I**	12	весь	12	에서	**В**
13	**він**	13	**that**	13	это	13	면	**Якщо**
14	**а**	14	**you**	14	как	14	한	**Один**
15	**це**	15	**he**	15	она	15	아/어서	**але**

* – корінь слова "робити";
** – корінь слова "бути";
*** –корінь слова "ставати".

Рис. 2

Отже, на основі презентабельного корпусу, що відображає сучасну письмову норму корейської мови, було укладено перший частотний морфемник корейської мови. На основі виведеного списку словоформ опираючись на зведений реєстр граматичних структур було укладено частотні списки граматичних та лексичних морфем які мають особливу цінність для досягнення дидактичних цілей та вирішення багатьох лінгвістичних задач. Постає питання про укладання підручника корейської мови на принципі частотності, використовуючи як базу частотний словник.

В статье расматриваются этапы составления частотного морфемника корейского языка. Проведен анализ корпуса, на материале которого он был составлен. Рассмотрена роль и место частотного словаря в современной лингводидактике.
Ключевые слова: *частотный словарь, корпус, корейский язык.*

Article observes all the stages of composing of the frequency dictionnary of the Korean language, its role in modern linguistic didactic in comparison with other languages and the analysis of the corpus on which the frequency dictionary was made.
Key words: *frequency dictionary, corpus, Korean language.*

Література:

1. *Засоріна Л. Н.* Частотный словарь русского языка /Л.Н. Засоріна. – М., 1977.
2. *Николова Ц.* Речник за българска разговорна реч / Ц. Николова – София, 1987.
3. *Рудий Б.А.* Text analyzer / Б.А. Рудий. – 2005. – E-product.

4. *Тищенко К.Н.* Лингвостатистческие законы и содержание обучения языку / К.Н. Тищенко // Вестник Киевского университета. Романо-герм. филол. – К., 1985. – Вып.19. – С. 3–8.

5. *Штейнфельдт Э.А.* Частотный словарь современного русского литературного языка / Э.А. Штейнфельдт. – Таллин, 1963.

6. *Josselson H.H.* The Russian Word Count and Frequency Analysis of Grammatical Categories / H.H. Josselson. – Detroit, 1953.

7. *Saukkonen P.* Finnish freq dictionary / *Pauli* Saukkonen. L., 1979.

8. *Thorndike E. L* The Teacher's Word book of 30000 Words / E.L. Thorndike, I. Lorge. – N.Y., 1972.

9. Електронний ресурс: http://blog.daum.net/youpd.

10. Електронний ресурс: http://www.cine21.com/Article.

11. Електронний ресурс: http://news.joins.com.

12. Електронний ресурс: http://www.report0u.com .

13. Електронний ресурс: http://www.yesform.com.

14. Електронний ресурс: http://monthly.chosun.com.

15. Електронний ресурс: http://www.sdt.or.kr.

16. Електронний ресурс: http://www.newsva.co.kr.

17. Електронний ресурс: http://womansense.ismg.co.kr.

18. Електронний ресурс: http://www.koreakidnews.org.