

**ВСТУП ДО
ПРИКЛАДНОЇ ЛІНГВІСТИКИ:
КОРПУСНА ЛІНГВІСТИКА І
ВИВЧЕННЯ МОВ**

Повторення матеріалу лекції 2 **ЛЕКСИКОГРАФІЯ**

Теоретична
Загальна

Прикладна
Спеціальна

Історія укладання словників:

глоси > тлумачні > національні > глобальні
рукописні > друковані > електронні

Типи словників

Основні проблеми лексикографії

КОРПУСНА ЛІНГВІСТИКА : ПОНЯТТЯ «КОРПУС»

- корпусна лінгвістика (*corpus linguistics; corpus linguist*)
- комп'ютерна лексикографія (*computational lexicography*)

a corpus is a large and principled collection of natural texts.

(Biber, Conrad, & Reppen, 1998, p. 12)

A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research.

(Sinclair, 2005)

A corpus is a collection of (1) machine readable (2) authentic texts (including transcripts of spoken data) which is (3) sampled to be (4) representative of a particular language or language variety.

(McEnery, Xiao, & Tono, 2006, p. 5)

ПРОБЛЕМИ СТВОРЕННЯ ЕЛЕКТРОННИХ СЛОВНИКІВ НА ОСНОВІ ДРУКОВАНИХ

- Ідеалізація слів і тенденція до науковості визначень (*spider “member of Arachnidae, class Aranea”*)
- Ігнорування динаміки лексикону (неологізм *appendicitis* – виключено зі словника Murray (1870))
- Невизначеність щодо власних назв (*Shakespeare, England* vs *Persil, Pepsi*)
- Складність диференціації значень (*to pour*)
- Ілюзорність меж слів та складність їх розмужування (*of course, insofar as, fire truck*)



Завдання :

Реструктурація словників у корпусний формат

- *Розгляньте корпуси та визначте:*
- Який матеріал вони містять.
- Що слугувало джерелами матеріалу.
- Яким чином матеріал організований.
- Як можна застосувати корпусні дані для дослідження або вивчення мови.



КОРПУСИ АНГЛІЙСЬКОЇ МОВИ

COCA Corpus of Contemporary American English

<https://www.english-corpora.org/coca/>

The Student-Transcribed Corpus of Spoken American English

<https://www.spokencorpus.org>

Та інші <https://www.english-corpora.org>