**Тема 7.** Корпус текстів природної мови. Корпусна розмітка.

1. Preliminary work on creating a corpus: problems of representativeness, selection of sources, external and internal selection criteria.
2. File normalization.
3. Graphematic analysis.
4. Parallel corpora.
5. Alignment problem.
6. Standardization in corpus linguistics.
7. Linguistic means of presenting parsed texts.

**Завдання для самостійної роботи**

Самостійна робота 4

Перегляд та коментування відео «Corpus linguistics for language lovers. - Ondřej Matuška | PG 2019».

Режим доступу: <https://www.youtube.com/watch?v=PtKMjj1yaCY>

**Індивідуальні завдання**

Підготувати усне повідомлення на одну із запропонованих тем:

1. Конкорданс у системі вивчення значення слова.
2. Використання конкордансів та корпусних технологій у сучасних лінгвістичних дослідженнях.

**Рекомендована література**

***Базова література:***

1. Biber D., Reppen R. English corpus linguistics. Cambridge University Press. 2015. 642 p.

2. Lange C., Leuckert S. Corpus Linguistics for World Englishes: A Guide for Research, New-York: Routledge, 2020. 236 p.

3. Rühlemann Ch. Corpus Linguistics for Pragmatics: A Guide for Research – New York, London : Routledge, Taylor &amp; Francis Group, 2019. 206 p.

4. Stefanowitsch, A. Corpus linguistics: a guide to the methodology. Berlin: Language Science, 2020. 490 р.

***Додаткова:***

1. Волошин В.Г. Комп’ютерна лінгвістика: Навч. посібник. Суми: Університетська книга, 2004.

2. Дарчук Н. П. Комп’ютерна лінгвістика (автоматичне опрацювання тексту): підручник. К.: Видавничо-поліграфічний центр “Київський університет”, 2008.

3. Карпіловська Є. А. Вступ до комп’ютерної лінгвістики. Донецьк: Юго- Восток, 2003.

4. Лук’янчук С. Комп’ютерна модель парадигматичних класів дієслів // Українське мовознавство. 2000. Вип. 22. С. 82-85.

5. Пещак М. М. Нариси з комп’ютерної лінгвістики. Ужгород: Видавництво закарпаття, 1999. 200 с.

6. Andrushenko, O. Iu. Lancsbox software options for the prospective investigation of the multilingual corpus for European studies. Вісник КНЛУ. Серія Філологія, 2023. Вип. 26 (1). С. 16–32.

7. Garcia Fernandes, L. Sources and Steps of Corpus Lemmatization. Revista Española de Lingüística Aplicada/Spanish Journal of Applied Linguistics. (2020). Vol. 33:2. Р. 416–442

8. Geyken, A. &amp; Gloning, T. (2015). A living text archive of 15th-19th-century

German: Corpus strategies, technology, organization. In J. Gippert &amp; R. Gehrke (Eds.), Historical corpora. Challenges and perspectives (pp. 165–179). Tübingen: Narr.

9. Lüdeling, A., J. Ritz, M. Stede, A. Zeldes. Corpus linguistics and information structure research, in: C. Fery, Sh. Ishihara (Eds.), The Oxford Handbook of Information Structure, OUP, Oxford, 2016, З. 599–620.

10. McEnery, T. &amp; Hardie, A. Corpus Linguistics. New-York: Routledge. 2015. 11. Toupin F., Gatelais S., Sasu I. (Eds.). Studies in Linguistic Variation and Change 3. Corpus Based Research in English Syntax and Lexis. Cambridge: Cambridge Scholar Publishing, 2020. 214 p.

12. Weisser M. Practical Corpus Linguistics: An Introduction to Corpus-Based Language Analysis. Wiley, Blackwell, 2016. 287 p.

***Інформаційні ресурси:***

1. Big data and Language 1 <https://www.coursera.org/learn/big-data-language-1>

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

3. https://www.ice-corpora.uzh.ch/en.html

4. https://childes.talkbank.org