



UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

A Corpus-Based Approach to Interpreting Studies: Corpus of Political Speeches in Ukrainian-English Combination

Anastasiia Novoshytska

Solidarity with Ukraine – 4EU+ for Ukraine II / International
Conference „European Narratives and Identity: Ukraine – A New
Perspective“, Paris, 25–26.06.2024

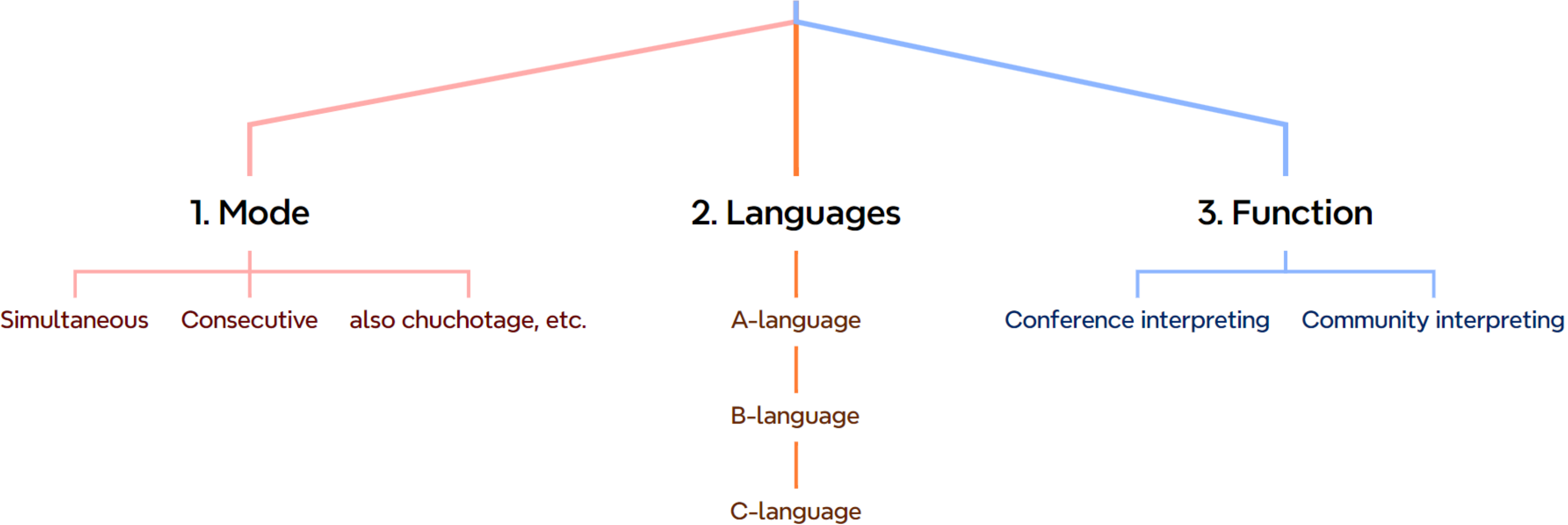
Agenda

1. What is interpreting?
2. The role of Ukrainian interpreters in today's world
3. Problems and opportunities
4. Corpus approach

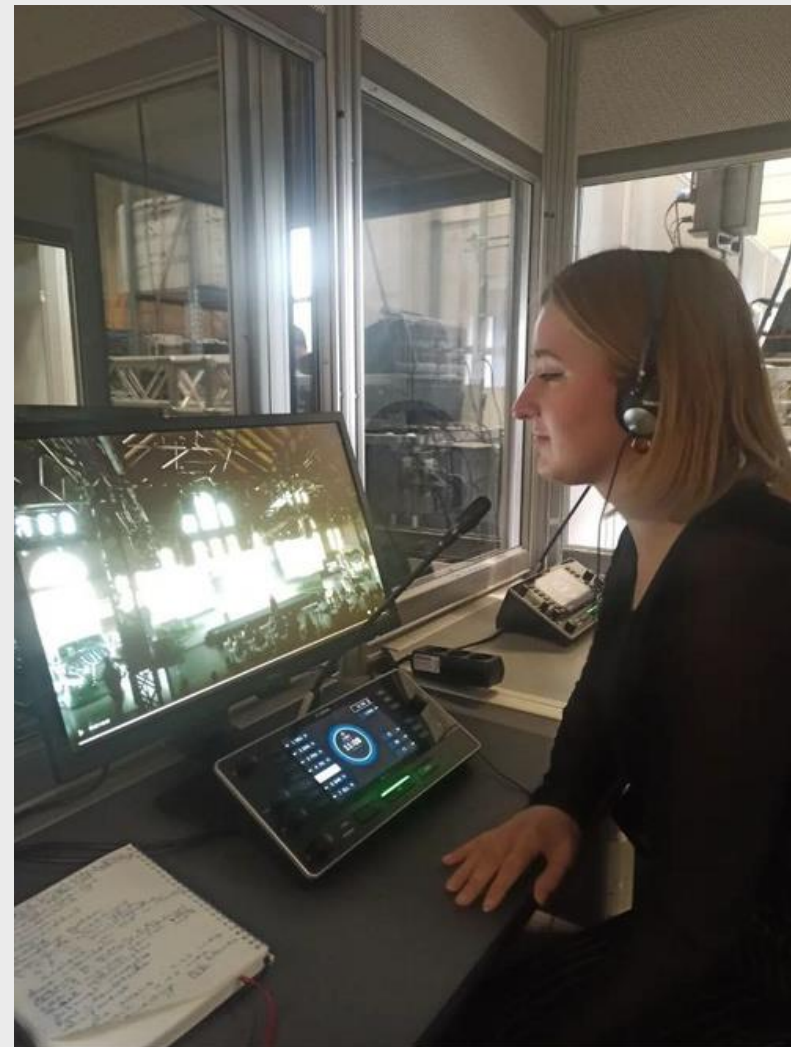
What is interpreting?

The International Association of Conference Interpreters (AIIIC) defines interpreting as the practice of conveying the meaning of a speaker's message orally and in another language to listeners who would not otherwise understand.

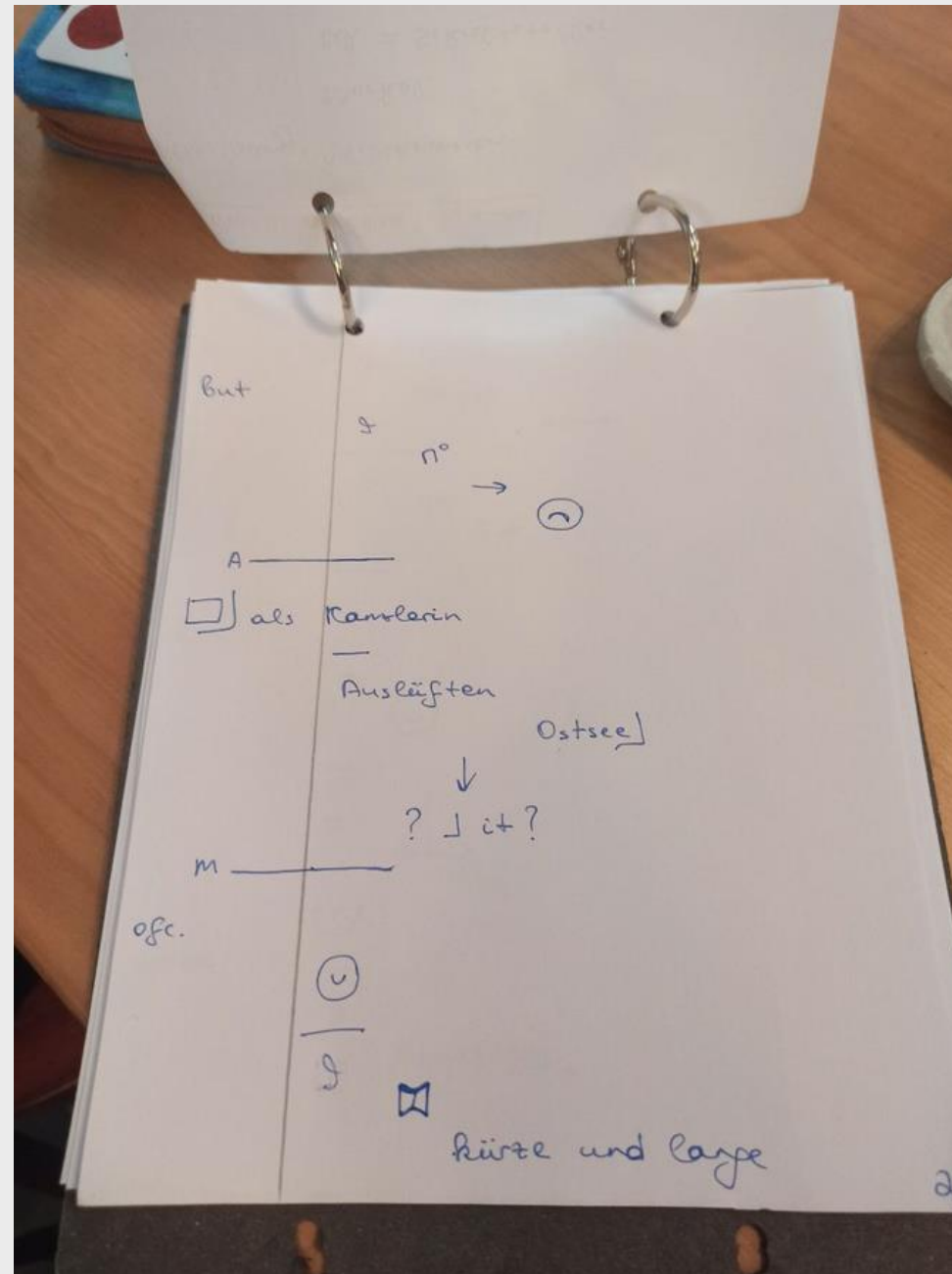
Interpreting



Simultaneous interpreting is a mode of interpreting in which the speaker makes a speech and the interpreter reformulates the speech into a language his audience understands at the same time (or simultaneously) (SCIC).



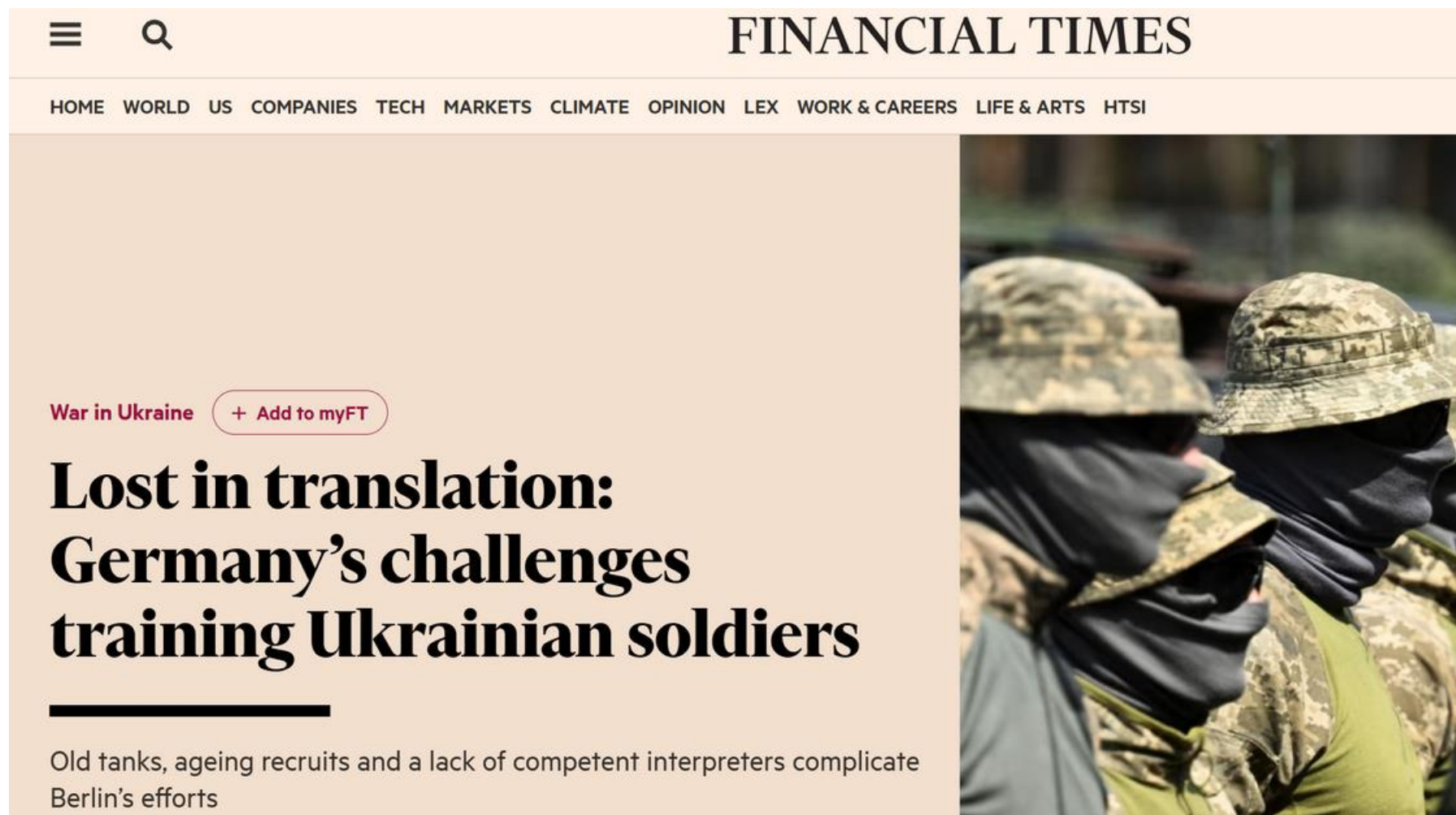
Consecutive interpreting involves the interpreter taking notes while the speaker talks, then reproducing the speech for the audience. The interpreter usually stands or sits close to the speaker, takes notes with a pad and pen, and interprets only when it's their turn to speak (SCIC).



2. The role of Ukrainian interpreters in today's world due to...

1. The Russian aggression and following consequences on the whole world in all possible areas

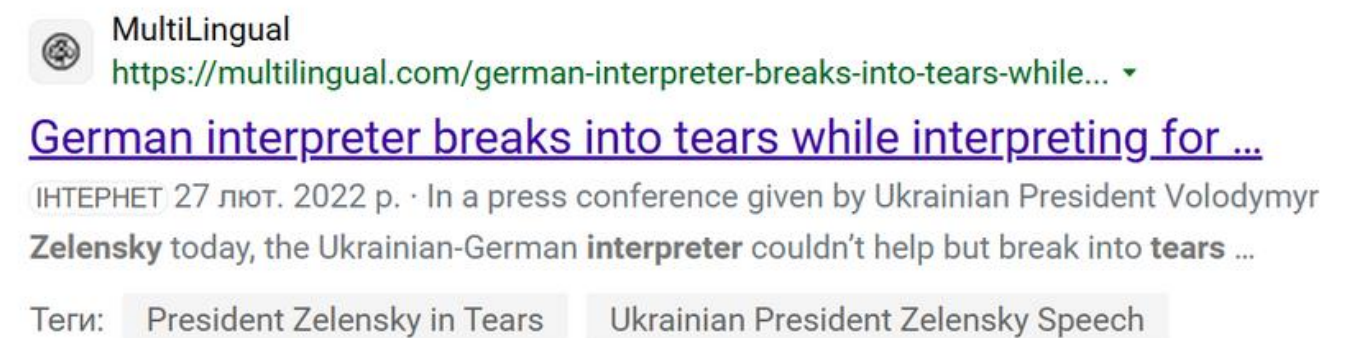
2. EU-candidacy



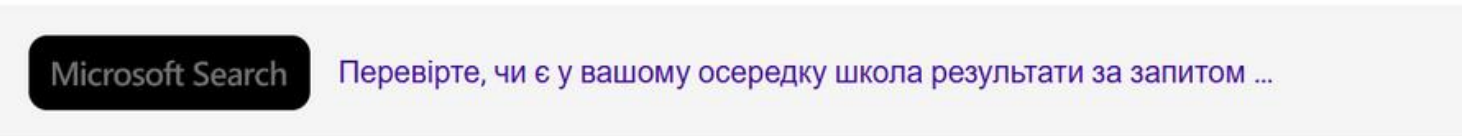
The screenshot shows the Financial Times website. The main headline is "Lost in translation: Germany's challenges training Ukrainian soldiers". A sub-headline reads "Old tanks, ageing recruits and a lack of competent interpreters complicate Berlin's efforts". The article is categorized under "War in Ukraine". To the right of the text is a photograph of two soldiers in camouflage gear and balaclavas.



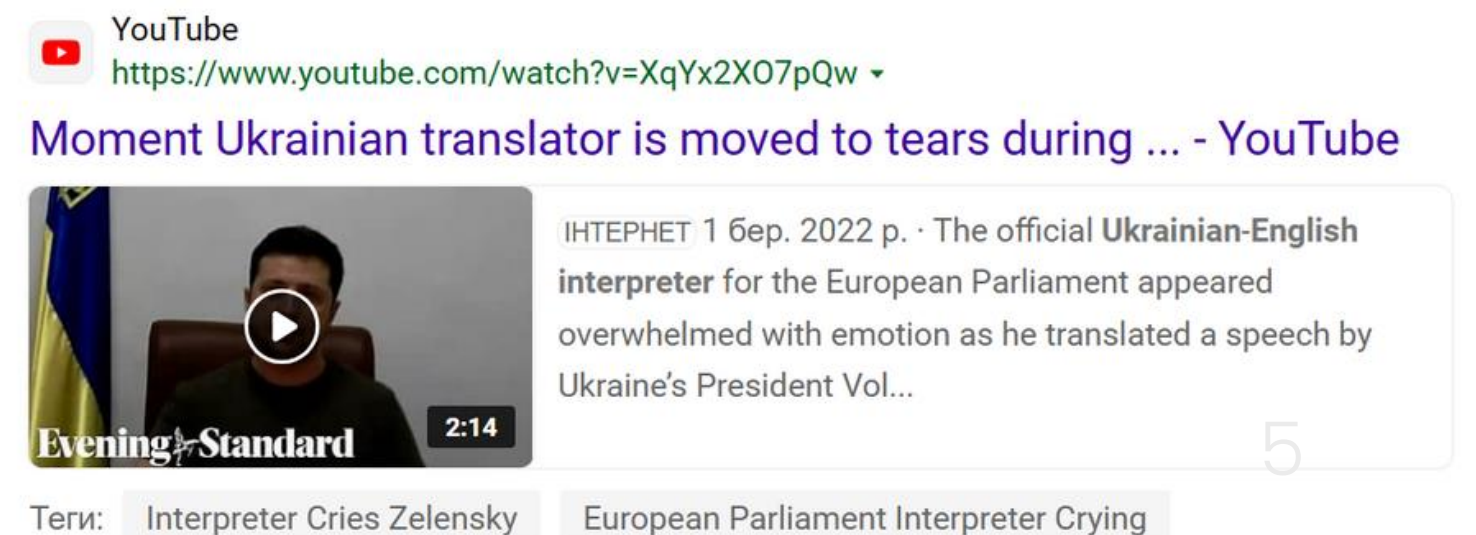
This is a search result snippet from Yahoo. The article title is "Veteran interpreter breaks down in tears after Zelensky remarks". The snippet includes the date "ІНТЕРНЕТ 27 лют. 2022 р." and a brief description: "A veteran interpreter broke down in tears on Sunday while translating Ukrainian President Volodymyr Zelensky's speech for a German news...". It also lists tags: "President Zelensky in Tears" and "Interpreter Breaks Down Zelensky".



This is a search result snippet from MultiLingual. The article title is "German interpreter breaks into tears while interpreting for...". The snippet includes the date "ІНТЕРНЕТ 27 лют. 2022 р." and a brief description: "In a press conference given by Ukrainian President Volodymyr Zelensky today, the Ukrainian-German interpreter couldn't help but break into tears...". It also lists tags: "President Zelensky in Tears" and "Ukrainian President Zelensky Speech".



The screenshot shows a Microsoft Search bar. The search query is in Ukrainian: "Перевірте, чи є у вашому осередку школа результати за запитом...".

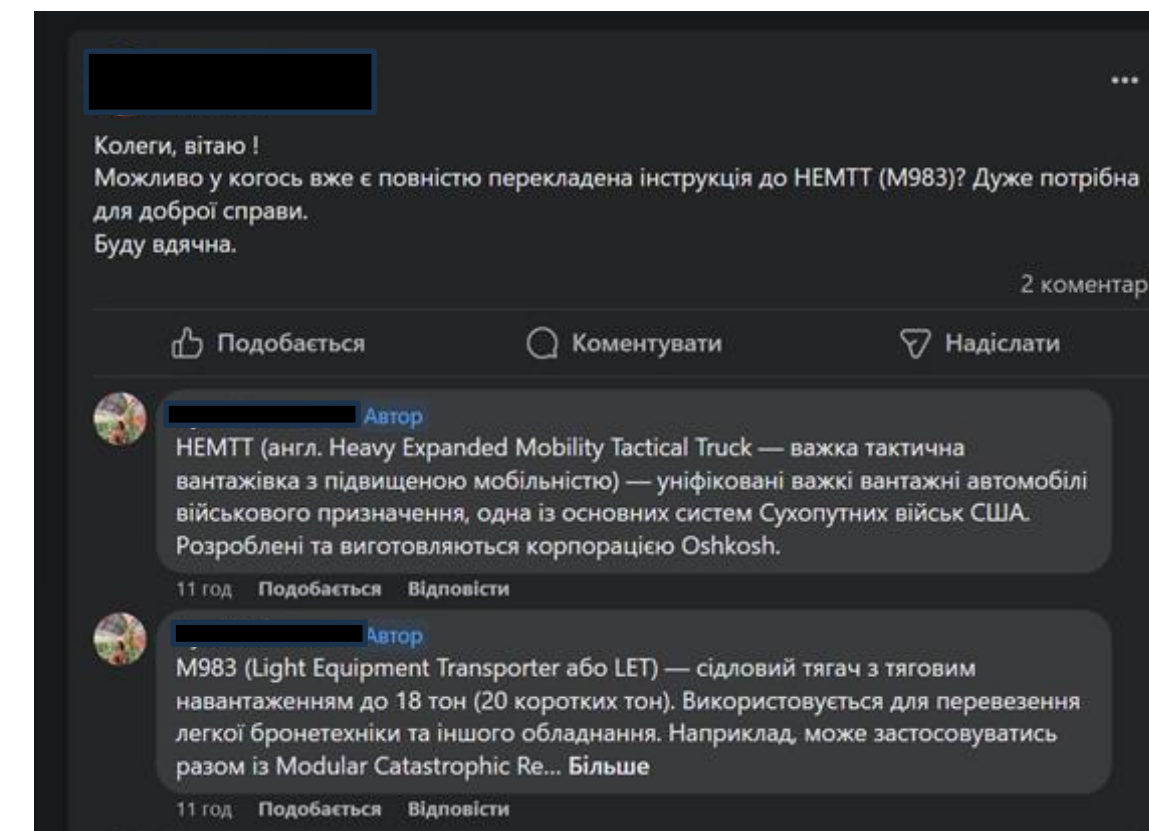


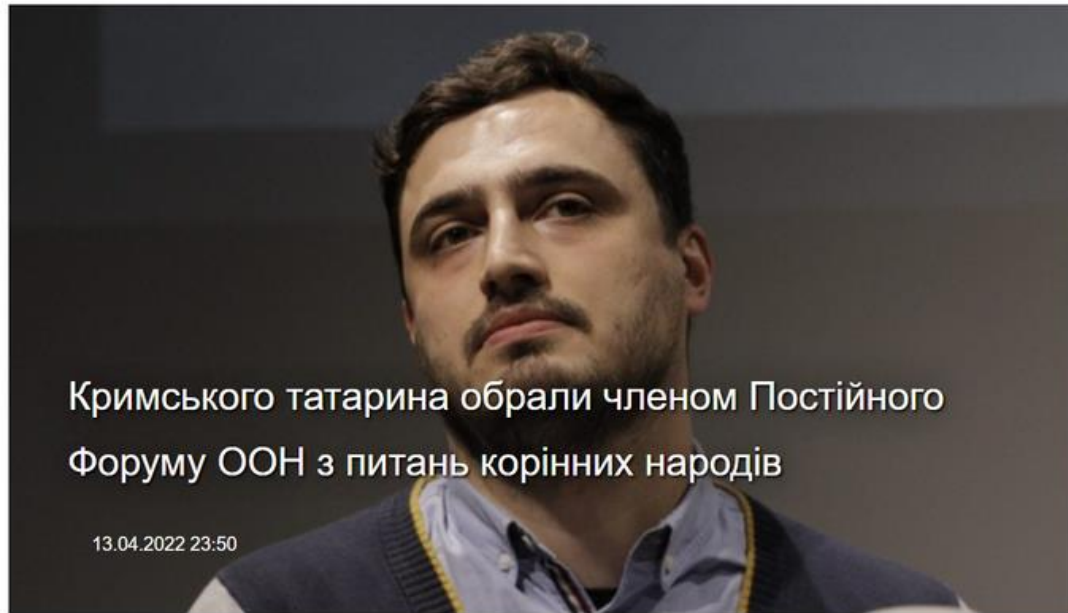
This is a search result snippet from YouTube. The video title is "Moment Ukrainian translator is moved to tears during...". The snippet includes the date "ІНТЕРНЕТ 1 бер. 2022 р." and a brief description: "The official Ukrainian-English interpreter for the European Parliament appeared overwhelmed with emotion as he translated a speech by Ukraine's President Vol...". It also lists tags: "Interpreter Cries Zelensky" and "European Parliament Interpreter Crying".

**Problems and
opportunities/ideas**

Problems

- Lack of professionals
- Often low quality of interpretation
- Lack of materials
Not qualitative materials for preparation and development of interpreters
- Lack of study programs
- Necessary databases have been only emerging but they are needed already now: military affairs, medicine, psychology
- Market, wage, competition





Членом Постійного Форуму ООН з питань корінних народів обрано представника Української держави, кримського татарина Сулеймана Мамутова.

Кримськотатарське питання на Експертному механізмі ООН

18 Липня, 2023 11:12

Як повідомив 17 липня член Меджлісу кримськотатарського народу Ескендер Барієв він та представниця Кримськотатарського ресурсного центру Людмила Короткіх розпочали роботу на 16 щорічній сесії Експертного механізму ООН з прав корінних народів.

But at the same time, the Ukrainian interpretation sector has been developing rapidly.

- Ukrainian is becoming significant in the EU. These and many other factors lead to...
- First study courses on interpreting in Ukraine and abroad.
- Grants and programs from the EU for Ukrainian interpreters, interpreting trainers, and programs.
- Ukrainian language courses are offered in various places around the world.
- More and more interpreters are learning Ukrainian.
- More and more conferences are being held in Ukrainian.
- Emerging need: Interpreting from/into the Crimean Tatar language.

Corpus approach

The outline of corpus creation:

For this stage of the study, I have developed an electronic parallel corpus comprising transcripts of original speeches and their interpreted versions in Ukrainian and English, delivered during the ninth term of the European Parliament from 2022 to 2023. The transcripts were obtained from the official website of the European Parliament. To compile, manage, and analyze the corpus effectively, I utilized commercial corpus management and corpus query software, SkethEngine¹. Although the selected speeches were interpreted simultaneously, the corpus and study results are equally applicable to the consecutive mode. The following are the steps I undertook to create a corpus:

1. *Corpus Design*

1.1 Speech Criteria

- Language combination: English ↔ Ukrainian
- Mode: Simultaneous interpretation
- Topic: War in Ukraine, Politics, EU
- Source: Database of plenary sessions of the European Parliament

1.2 Defining Scope and Size




2. Collecting Speeches

3. *Transcription through Automatic Speech Recognition (ASR)*

4. *Alignment*

5. *Didactic Materials (metadata, symbols, scripts)*

The metadata for speeches in the thesis is partly developed based on parameters from the 'Speech repository,' a collection of monolingual recordings that includes excerpts from conferences or other public meetings, as well as speeches specifically designed for the training of conference interpreting students and practicing interpreters. The collection has been developed by the Directorate-General (SCIC), the Commission's interpreting service and multilingual conference organiser (Directorate-General SCIC)². The recordings on the "Speech Repository" are organized by the following parameters:

General information	Link: Sitting of 14-12-2022 Plenary European Parliament (europa.eu) Name: Award of the Sakharov Prize (formal sitting) Date: 14.12.2022 Speaker: Volodymyr Zelensky Place: European Parliament, Brussels Occasion (if any): Style: formal	
Description	President of Ukraine, addresses an audience, expressing gratitude for their support in Ukraine's fight for freedom. He condemns Russia's aggression, emphasizes the importance of Ukraine to Europe, and calls for justice for the victims through international cooperation.	
Terminology	Sakharov Prize International Criminal Court, UN General Assembly, Tribunal, European Parliament resolution Mariupol, Volnovakha, Bakhmut, Donbas Holodomor	
Technical Details	Language: uk (original), en (target) Duration: 11:00 Level: Intermediate Mode: Simultaneous Use: Simultaneous, consecutive Type: Real-life speech Domains: Politics	
Symbols	свобода  війна  нищити, руйнувати, руїни, розбите 	
Script		
Original speech		Interpreted speech
Дякую, Роберто! Шановна пані Президент, дорога Роберто! Дорогі друзі! Я вітаю вас від імені всіх наших людей – велика честь. Я вітаю вас від усіх тих, які воюють, які працюють заради України й заради свободи. Заради того, без чого неможливо уявити всіх нас, життя Європи. Неможливо уявити без України і без свободи. І не лише тому, що це		Madam President, thank you, dear Roberta, dear friends. I greet you on behalf of all our people. It is a great honour. I greet you on behalf of those who are fighting, who are working for the sake of Ukraine and for the sake of freedom – for the sake of something without which it is impossible to imagine all of us, life, Europe. It is impossible to imagine Europe without Ukraine and without

SPEECH REPOSITORY

Speeches | Speech Reposi x +

speech-repository.webcloud.ec.europa.eu/search-speeches?keyword=&collection=&speaker=&language=221...

Type
- Any -

Refine results

Clear filters

EPRS - Speeches that have made Europe: Volodymyr Zelenskyy (2022)

Use	Level	Type	Domains	Duration
Simultaneous	Advanced / Test-type	Real-Life	Geopolitics	13:37

public (uk) Українська

Award of the Sakharov Prize 2022: statement by Volodymyr Zelenskyy

Use	Level	Type	Domains	Duration
Simultaneous	Advanced / Test-type	Real-Life	Culture	11:23

public (uk) Українська

Ruslan Stefanchuk urges EU to grant Ukraine candidate status

Use	Level	Type	Domains	Duration
Simultaneous	Advanced / Test-type	Real-Life	Geopolitics	14:20

public (uk) Українська

Балканський досвід вступу до ЄС

Use	Level	Type	Domains	Duration
Simultaneous	Very advanced	Pedagogical	Enlargement Politics	16:04

Щедрик: невідома історія відомої пісні x +

https://speech-repository.webcloud.ec.europa.eu/speech/щедрик-невідома-історія-відомої-пісні

Speech Repository

Search Speeches About Tutorial SCICRec FAQ

Speech Repository > Speech > Щедрик: невідома історія відомої пісні

Щедрик: невідома історія відомої пісні


public

Place: Brussels

Date: 22/02/2023 - 11:33

Description: Історія української народної пісні «Щедрик» та її значення для української незалежності.

Send us your feedback



Speech Details

Language: (uk) Українська

Speech number: 32292

COLLECTION OF VIDEOS

Промови

Пошук

Файл Основне Вставлення Малювання Макет сторінки Формули Дані Подання Автоматизація Acrobat

Автозбереження Зберегти Скасувати Скасувати Шрифт: Times New R Очистити Об'єднати та розташувати в центрі Переносити


B12

Plenary Sessions: Debates and Videos Plenary European Parliament (europa.eu)		
Date	Speech	1. Video 2. Script
8.6.22	Speech by Ruslan Stefanchuk, Speaker of the Verkhovna Rada	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220608-11:43:16&playerEndTime=20220608-12:02:33# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-06-08-ITM-007_EN.html
2.2.23	Preparation of the EU-Ukraine Summit (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20230202-08:32:43&playerEndTime=20230202-10:08:40# https://www.europarl.europa.eu/doceo/document/CRE-9-2023-02-02-ITM-002_EN.html
1.3.22	Russian aggression against Ukraine (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220301-12:34:38&playerEndTime=20220301-15:01:20# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-03-01-ITM-009_EN.html
1.3.22	Russian aggression against Ukraine (continuation of debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220301-15:06:03&playerEndTime=20220301-15:50:09# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-03-01-ITM-011_EN.html
9.2.23	Formal sitting - Address by Volodymyr Zelenskyy, President of Ukraine	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20230209-11:00:28&playerEndTime=20230209-11:28:12# https://www.europarl.europa.eu/doceo/document/CRE-9-2023-02-09-ITM-004_EN.html
15.6.23	Ukraine's accession to the Convention of 2 July 1919 on the Recognition and Enforcement of Foreign Judgments in Civil or Commercial Matters (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20230615-09:00:38&playerEndTime=20230615-09:15:06# https://www.europarl.europa.eu/doceo/document/CRE-9-2023-06-15-ITM-003_EN.html
19.5.22	The fight against impunity for war crimes in Ukraine (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220519-09:11:51&playerEndTime=20220519-10:16:49# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-05-19-ITM-003_EN.html
5.10.22	Russia's escalation of its war of aggression against Ukraine (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20221005-09:01:22&playerEndTime=20221005-12:10:21# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-10-05-ITM-002_EN.html
19.10.22	Lukashenko regime's active role in the war against Ukraine (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20221019-18:53:46&playerEndTime=20221019-19:57:44# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-10-19-ITM-017_EN.html
13.9.22	Question Time (VPC/HR) The state of play of the war in Ukraine	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220913-15:05:56&playerEndTime=20220913-16:20:53# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-09-13-ITM-013_EN.html
5.4.22	EU Protection of children and young people fleeing the war against Ukraine (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220405-09:07:39&playerEndTime=20220405-11:33:48# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-04-05-ITM-003_EN.html
17.1.23	The establishment of a tribunal on the crime of aggression against Ukraine (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20230117-20:52:01&playerEndTime=20230117-21:51:49# https://www.europarl.europa.eu/doceo/document/CRE-9-2023-01-17-ITM-016_EN.html
18.10.22	Impact of Russian invasion of Ukraine on migration flows to the EU (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20221018-19:54:08&playerEndTime=20221018-21:06:05# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-10-18-ITM-016_EN.html
19.4.23	Children forcibly deported from Ukraine and the ICC arrest warrant for Vladimir Putin (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20230419-09:02:57&playerEndTime=20230419-10:25:01# https://www.europarl.europa.eu/doceo/document/CRE-9-2023-04-19-ITM-005_EN.html
22.6.22	Draft amending budget No 3/2022: financing reception costs of people fleeing Ukraine (short presentation)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220622-22:26:35&playerEndTime=20220622-22:40:05# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-06-22-ITM-024_EN.html
8.3.22	The deterioration of the situation of refugees as a consequence of the Russian aggression against Ukraine (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220308-15:01:16&playerEndTime=20220308-18:33:23# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-03-08-ITM-009_EN.html
23.6.22	Draft amending budget No 3/2022: financing reception costs of people fleeing Ukraine (A9-0181/2022 - Karlo Ressler)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220623-12:28:49&playerEndTime=20220623-12:31:33# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-06-23-ITM-009-03_EN.html
4.5.22	The social and economic consequences for the EU of the Russian war in Ukraine - reinforcing the EU's capacity to act (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20220504-08:36:09&playerEndTime=20220504-10:48:02# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-05-04-ITM-004_EN.html
18.10.22	Keep the bills down: social and economic consequences of the war in Ukraine and the introduction of a windfall tax (debate)	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20221018-09:02:03&playerEndTime=20221018-10:50:26# https://www.europarl.europa.eu/doceo/document/CRE-9-2022-10-18-ITM-002_EN.html
17.1.23	The EU's response to the appalling attack against civilians in Dnipro : strengthening sanctions against the Putin regime and military support to Ukraine	https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20230117-20:05:13&playerEndTime=20230117-20:52:01#


Завдання Промови 16.02.23 01.03.22 01.03.22_(2) 09.03.22 06.04.22 22.06.22 05.04.22 19.05.22 19.5.22_(2) 08.06

PLENARY SESSIONS

← ↻ 🔒 https://www.europarl.europa.eu/plenary/en/vod.html?mode=chapter&vodLanguage=EN&playerStartTime=20230613-... 🔍 🔊 ☆ ⚙️ 📄 📌 🗨️ ⋮ 🌐

 European Parliament


Sitting of 13-06-2023



2023-06-13 • 09:01 - 10:29
Plenary session
Plenary session
▶ 🔊 10:25:27 / 10:29:31

🕒 09:01:40 - Start of speaking time

- 10:18:26 Michal WIEZIK
- 10:19:24 Roberts ZĪLE
- 10:19:25 Mónica Silvana GONZÁLEZ
- 10:20:31 Roberts ZĪLE
- 10:20:38 Billy KELLEHER
- 10:21:43 Roberts ZĪLE
- 10:21:51 Mick WALLACE
- 10:22:56 Roberts ZĪLE
- 10:23:01 Juozas OLEKAS
- 10:24:08 Roberts ZĪLE
- 10:24:13 Georgios KYRTSOS
- 10:25:04 Roberts ZĪLE
- 10:25:20 Valdis DOMBROVSKIS

09:01:40 ▶ 10:29:31
Humanitarian and environmental consequences of the destruction of the Nova Kakhovka dam - Sustainable reconstruction and integration of Ukraine into the Euro-Atlantic community (debate)

 **Current speaker**


Valdis DOMBROVSKIS
Λετονία

Duration ▶ 00:01:26 | Start ▶ 10:25:20 | End ▶ 10:26:46

AUTOMATIC SPEECH RECOGNITION (ASR) FOR UKR

The screenshot shows a web browser displaying a GitHub repository page. The browser's address bar shows the URL: `https://github.com/iued-uni-heidelberg/corpusdev/blob/main/Transcribe_Audio_With_Whisper_Ukrainian_UK_UA_UA_UK...`. The repository name is `iued-uni-heidelberg / corpusdev`, which is public. The file being viewed is `corpusdev / Transcribe_Audio_With_Whisper_Ukrainian_UK_UA_UA_UK.ipynb`, created by `iued-uni-heidelberg` using Colaboratory, 3 weeks ago. The file has 474 lines (474 loc) and is 19.1 KB. The page features a 'README' section with a yellow star icon and a 'Documentation' section with a person icon. The README text states: "This is the companion Colab for the article 'How to transcribe your audio to text, for free (with SRTs/VTs!)'." and "This Colab shows how to use OpenAI's Whisper to transcribe audio and audiovisual files, and how to save that transcription as a plain text file or as a VTT/SRT caption file." The left sidebar shows a file tree with folders `docs` and `notebooks`, and files including `LICENSE`, `README.md`, `S104SentenceAlignmentHunalign...`, `TerminologyExtraction4EUplus_v...`, `TerminologyExtractionV01.ipynb`, `Transcribe_Audio_With_Whisper.i...`, `Transcribe_Audio_With_Whisper_...`, and `Untitled8.ipynb`.

SKETCH ENGINE

The screenshot shows the Sketch Engine dashboard for the corpus 'Промови., English'. The interface includes a search bar, a grid of analysis tools, a 'RECENTLY USED CORPORA' list, and a promotional banner for a Lexicom workshop.

SEARCH BAR: Промови., English

ANALYSIS TOOLS:

- Word Sketch: Collocations and word combinations
- Word Sketch Difference: Compare collocations of two words
- Thesaurus: Synonyms and similar words
- Concordance: Examples of use in context
- Parallel Concordance: Translation search
- Wordlist: Frequency list
- N-grams: Multiword expressions (MWEs)
- Keywords: Terminology extraction
- Trends: Diachronic analysis, neologisms
- Text type analysis: Statistics of the whole corpus
- OneClick Dictionary: Automatic dictionary drafting
- Bilingual terms: Bilingual terminology extraction

RECENTLY USED CORPORA:

Corpus Name	Language	Count	Action
Промови., English	English	1,509	Trash
Промови., Ukrainian	Ukrainian	1,239	Trash
myTest2	Ukrainian	647,694	Trash
16.03.04	Spanish	43,860	Trash

PROMOTIONAL BANNER:

<Lexicom/>
An intensive workshop in digital lexicography and lexical computing
Málaga, Spain, 16–20 September 2024
[REGISTRATION](#)

NAVIGATION: MY SEARCH HISTORY | ANNOTATIONS

EU-High-Level Scientific Conference Series

MuTra 2005 – Challenges of Multidimensional Translation : Conference Proceedings

Claudio Bendazzoli & Annalisa Sandrelli (Forlì, Bologna)¹

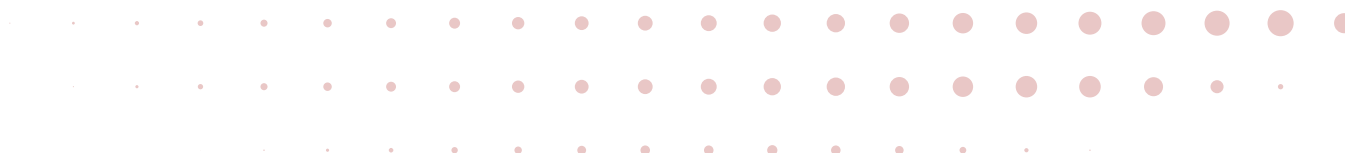
An Approach to Corpus-Based Interpreting Studies: Developing EPIC (European Parliament Interpreting Corpus)

Contents

- 1 Introduction
- 2 Methodological issues in interpreting research
- 3 The EPIC multimedia archive
- 4 The EPIC corpus
- 5 Research and pedagogical applications
- 6 Conclusions
- 7 References

Abstract

Empirical research on simultaneous interpreting is hampered by the problem of collecting sufficient material (recordings of source speeches and interpreted target speeches) for the testing of hypotheses and validation of existing theories. In other words, corpora have long been awaited in the field of Interpreting Studies. In January 2004 a research group on corpus-based interpreting studies was set up in the Department of Interdisciplinary Studies in Translation, Languages and Cultures (SITLeC) of the University of Bologna at Forlì, in order to create an electronic parallel



Useful links

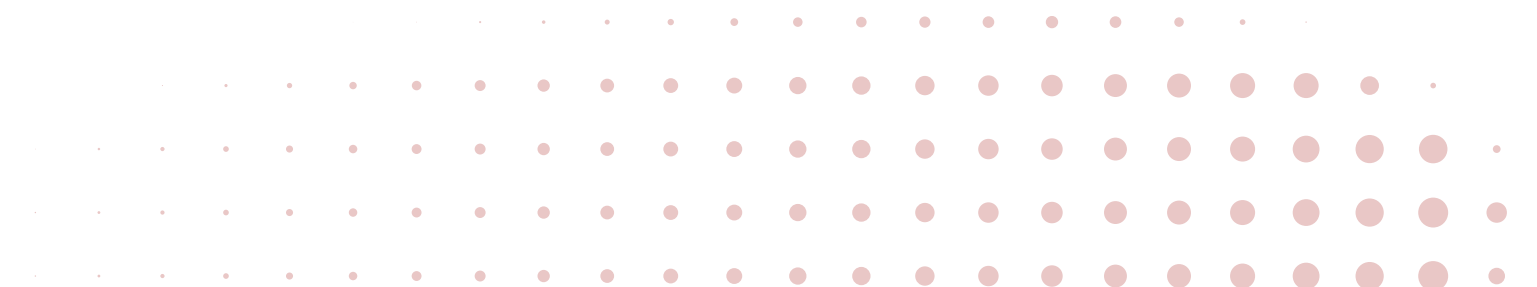
Bendazzoli, C. (2021). Corpus studies in conference interpreting. In M. Albl-Mikasa, & E. Tiselius (Eds.), *The Routledge Handbook of Conference Interpreting* (1 ed., p. 14). London: Routledge.

[doi:doi.org/10.4324/9780429297878](https://doi.org/10.4324/9780429297878)

Bowker, L., & Pearson, J. (2002). *Working with Specilized Language: A Practical Guide to Using Corpora*. New York: Routledge.

Corpas Pastor, G. (2016). *Corpus-based approaches to translation and interpreting*. Frankfurt am Main ; New York.

Setton, R. (2013). *Corpus-Based Interpreting Studies (CIS): Overview and Prospects*. In A. Kruger. London. Retrieved 10 19, 2023



**Thank you
for your attention!**

Anastasiia Novoshytska