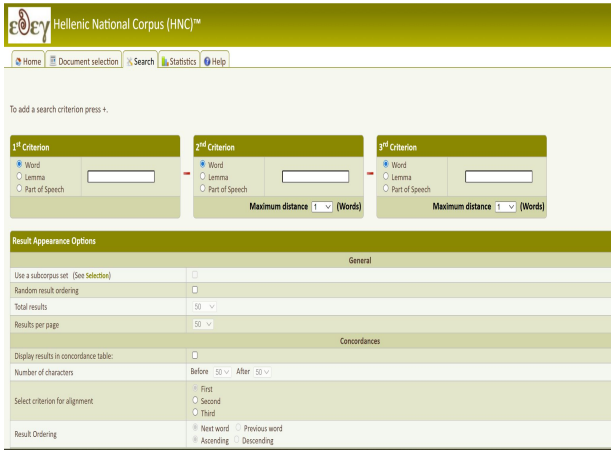
CMLC-12
Flash presentations of posters

(@LREC) Palma, Mallorca, 11.05.2026

Hellenic National Corpus: the current state

https://hnc.ilsp.gr/



- corpus of standard Modern Greek
- contains only written language
- from a variety of sources, genres and text types
- it is well-documented with metadata
- and automatically annotated (lemmatized and PoS tagged)
- it is accessible via a dedicated environment
- A small subset, automatically annotated and additionally manually corrected, named *Golden Part of Speech Tagged corpus* is available for download thorough CLARIN:EL

Hellenic National Corpus: the current state

The HNC access platform includes two modules: **backend** (responsible for data preparation, curation and storage) and **frontend** (responsible for the user interface and visualizations of the results)

Backend consists of	**Frontend caters for**
● user management environment ● text management environment ● metadata editor	● *search,* which produces concordances ● analysis and correlation of words or lemmas ● statistical information: word / lemma frequencies

Future Steps

The future steps for HNC fall into 5 main axes:

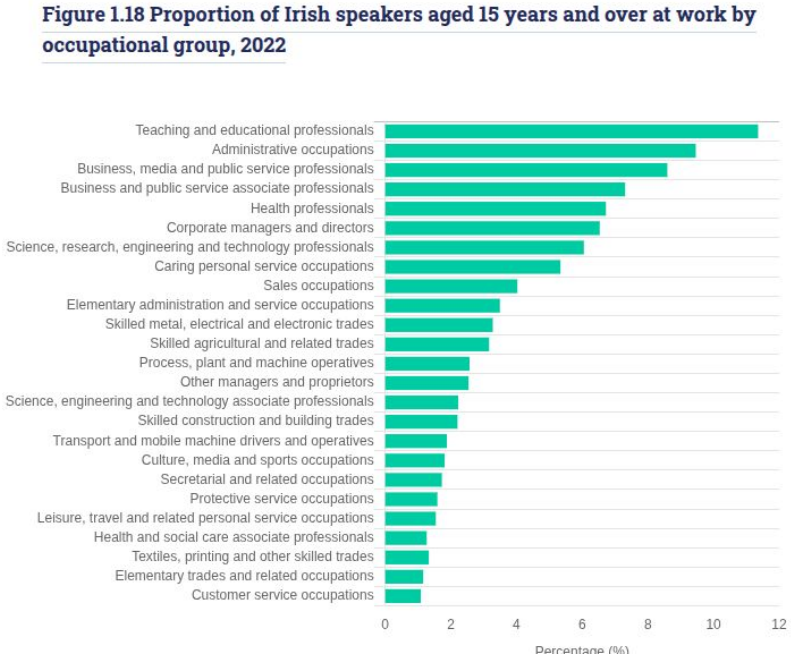
- content enrichment with older material (diachronic expansion)
- addition of dialectal material (geographical expansion)
- experimentation with existing LLMs for the annotation of standard Modern Greek texts
- fine-tuning of existing models for the processing of dialectal material
- distribution of more open-access material through CLARIN:EL

Corpas Náisiúnta na Gaeilge 2022-2029: A Project Overview

- During **Phase 1 of 3** of our project we compiled **4 corpora** for different purposes, one of which is the National Corpus.

While the national corpus is a balanced corpus, as a smaller language community we do suffer from some domain restriction

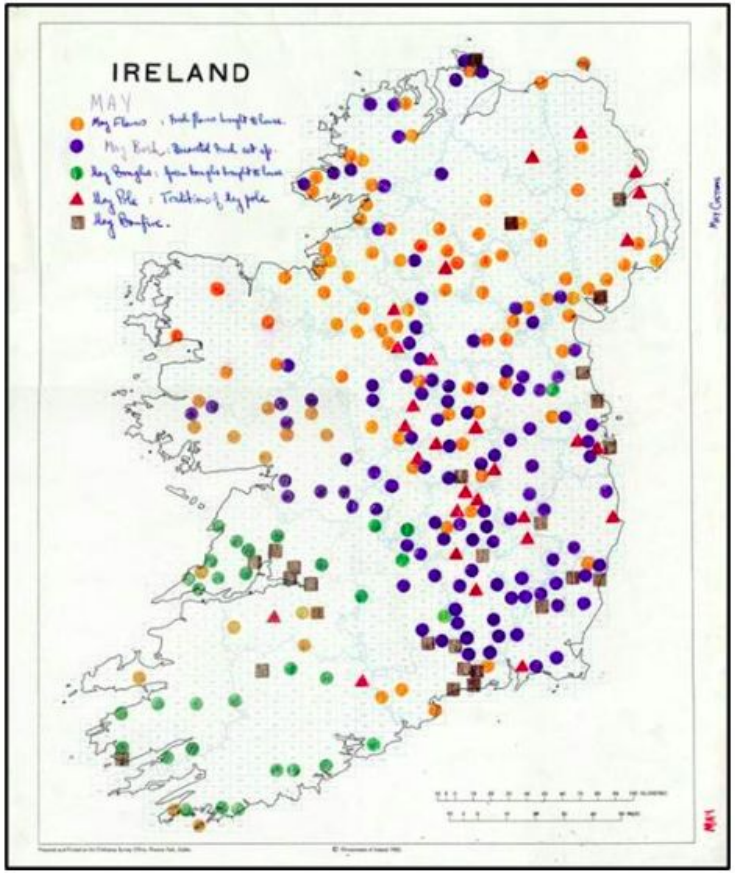
The use cases and needs of our community also vary slightly to those of other national corpora.



Corpas Náisiúnta na Gaeilge 2022-2029: A Project Overview

- **Phase 2 of 3** was a one-year extension to the initial funding phase, where we focused on the monitor corpus and on spoken data.
- **Phase 3 of 3** commenced in January of this year, extending to 2029. During this phase wel will add **parallel Irish-English legislative data**, top-ups to existing corpora, and the addition of **The Schools Collection** a UNESCO recognised folklore archive of stories written by school children in the early 1930s in Ireland.

Can you guess what challenges lay ahead!?



Currently available through a sister website: www.duchas.ie

School:
Fearainn an Choirce
(roll number 14532)

Location:
Fearann an Choirce, Co. Galway

Teacher:
Seán Ó Maoldhomnaigh

Browse

TITLES (97)

1. Ciste Óir
Transcribed

2. Ciste Óir
Transcribed

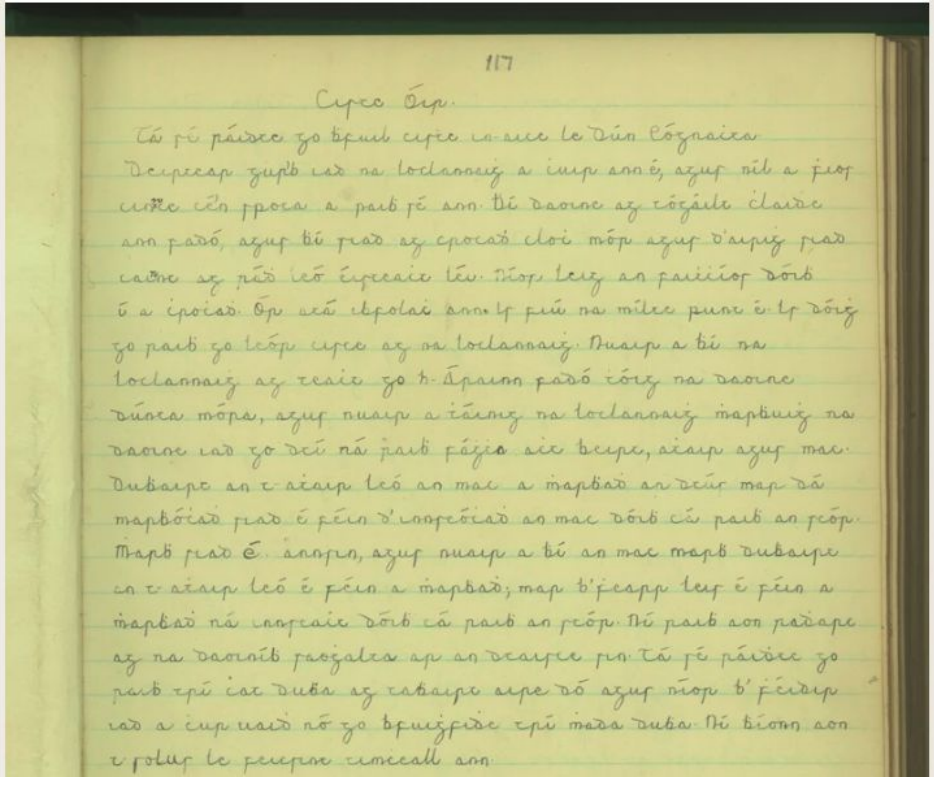
3. Ciste Óir
Transcribed

4. Ciste Óir
Transcribed

5. Ciste Óir
Transcribed

MODE: Magnify | Zoom

117 / 282



ON THIS PAGE

Ciste Óir

Share Post

Tá sé ráidte go bhfuil ciste in aice le Dún Eógnachta. Deirtear gur'b iad na loclannaigh a chuir ann é, agus níl a fhios cinnte cé'n spota a raibh sé ann. Bhí daoine ag tógáilt chlaidhe ann fadó, agus bhí siad ag crocadh cloch mór agus d'airigh siad cainnt ag rádh leo éisteacht lei . Níor leig an faithchios dóibh í a chrochadh .Ór atá ibfolach ann. Is fiú na mílte punt é .Ir dóigh go raibh go leór ciste ag an lochlannaigh . Nuair a bhí na loclannaigh ag teacht go h-Árainn fado thóig na daoine dúnta móra , agus nuair a tháinig na loclannaigh mharbhuigh na daoine iad go dti ná raibh fághtha acht beirt, athair agus mac. Dubhairt an t-athair leó an mac a mharbhadh an dtúr mar dá marbhóchadh siad é annsin agus nuair a bhí an mac marbh dubhairt an t- athair leó é féin a mharbhadh ; mar b'fhearr leis é féin a marbhadh ná innseacht dóibh cá raibh an stór . Ní raibh aon radharc ag an daoiníbh saoghalta ar an dtaisce sin . Tá sé ráidhte go raibh tri chat dubha ag tabhairt aire

General Regionally Annotated Corpus of Ukrainian (GRAC): Recent Developments and Future Plans

- The largest manually curated and annotated corpus of Ukrainian (2016–)
- >2 billion tokens · >800,000+ texts · >35,000 authors · 1816–2025
- Open volunteer project hosted at University of Jena
- 1,600+ students from 15+ Ukrainian universities contributed

FRIEDRICH-SCHILLER-UNIVERSITÄT JENA

ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ 1885 НТУ

UKRAINIAN CATHOLIC UNIVERSITY

ОДЕСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМ. І. І. МЕЧНИКОВА 1865

General Regionally Annotated Corpus of Ukrainian

Annotation layers

- Morphological annotation: TagText (based on the VESUM morphological dictionary by Andry Rysin, Vasyl Starko et al. 2005–)
  Accuracy: lemma 99.3% · POS 98.7%
- Semantic annotation: Ukrainian Semantic Lexicon (USL) (Vasyl Starko)
- Regional annotation: author's region of origin or place of publication — essential for tracking historical variation of Ukrainian

Derivative projects

- Syntactically annotated corpora: UD_Ukrainian_ParlaMint · Rada_Trees
- Parallel corpora: ParaRook · ParaFarm (multiple-translation corpus)
- Code-switching corpus: ParlMix-UA-RU

General Regionally Annotated Corpus of Ukrainian

Next version (GRAC.v.20) will add syntactic dependency annotation following the Universal Dependencies (UD) framework, using UDPipe2 (Straka et al., 2016–).

This will enable:

- precise search for grammatical constructions and syntactic patterns
- improved morphological disambiguation accuracy

Recent developments of the Bulgarian National Corpus

Svetla Koeva | svetla@dcl.bas.bg Ivelina Stoyanova | iva@dcl.bas.bg
Institute for Bulgarian Language, Bulgarian Academy of Sciences

Motivation and objectives

The main priorities of BulNC include:

- **diversity of data,**
- **extensive metadata description,**
- **linguistic integrity.**

Shift to NLP applications and the development of **IfGPT Dataset** – a large BulNC-based dataset with a special focus on LLMs fine-tuning.

IfGPT

Web Search Interface
Multimodal Corpus MIC-21
IfGPT Metadata Search Interface

https://ifgpt.dcl.bas.bg/

The work is part of the project *Infrastructure for Fine-tuning Pre-trained Large Language Models*, Grant Agreement No. ПВУ – 55 from 12.12.2024 /BG-RRP-2.017-0030-C01/.

Funded by the European Union NextGenerationEU
National Recovery and Resilience Plan of the Republic of Bulgaria

The British National Corpus 1994 to 2026

Megan Bushnell and Martin Wynne, Faculty of Linguistics, Philology and Phonetics, University of Oxford

- Created by a consortium of publishers and academics, funded by the Department of Trade and Industry;
- Original version completed in 1994 - oldest national reference corpus?
- 100m words: 90% writing, 10% speech (inc. 50% spontaneous conversation);
- No new texts added (some removed 1995 for copyright reasons);
- TEI XML version released 2003
- POS, lemma, structural tags frozen since 2003
- Custom BNC Licence
- Available online via various interfaces
- Available for download from the Oxford Text archive
- Lancaster University working on comparable corpora e.g. BNC2014

British National Corpus: 1994 to 2026

Megan Bushnell & Martin Wynne
Faculty of Linguistics, Philology and Phonetics
University of Oxford

BRITISH NATIONAL CORPUS

OXFORD TEXT ARCHIVE
<ota since="1976"/>

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a range of sources, designed to represent a wide cross-section of British English from the late 20th century. It is one of the first monolingual, synchronic, general, representative corpora of its size, and led the way for other national corpora. It was created by a consortium of academic partners and publishers, with funding from the Department of Trade and Industry in the UK. What are the lessons learned over the past thirty years regarding representativeness, modes of access, licensing, and managing the transition from a contemporary corpus to a historical one?

Versions of the BNC

- 1991-1994: Data collection and corpus building
- 1995: BNC1 first release (SGML) on CD
- 1995: BNC World (worldwide release) on CD bundled with Sara software
- 1998: BNC Sampler (2m word sample, speech and writing)
- 1998: BNC Handbook (textbook on using BNC with Sara)
- 1999: BNCWeb online interface launched (hosted by Lancaster University based on CQPweb)
- 2000: Available via BYU corpora online (now English-corpora.org)

CD cover of BNC XML, distributed by OUCS on behalf of the BNC Consortium.

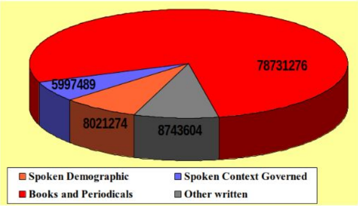
- 2003: BNC XML launched (TEI XML encoding, with improved POS tagging) on 2 CDs with Xaira software.

BNC BRITISH NATIONAL CORPUS 2014

Researchers at Lancaster University have produced spoken and written corpora comparable to the BNC with data up to 2014, available via the #LancsBox tool and CQPweb.

- 2003: BNC Baby (4m sample) released on CD
- 2007: Streaming audio added to BNCweb
- 2014: All BNC versions become available for download from OTA
- 2014: Lancaster build comparable BNC2014 corpora
- 2020: Available via Sketch Engine
- 2021: OTA funded as part of the national infrastructure again
- 2025: BNC World rebranded as BNC1994

The corpus composition




```
<u who="DCKPSUNK">
 <s n="36">
  <w c5="AV0" hw="so" pos="ADV">So </w>
  <w c5="VBZ" hw="be" pos="VERB">is </w>
  <w c5="PNP" hw="it" pos="PRON">it </w>
  <w c5="PRP" hw="like"
  pos="PREP">like</w>
  <c c5="PUN">, </c>
  <w c5="PNI" hw="somebody"
  pos="PRON">somebody </w>
  <w c5="VM0" hw="could" pos="VERB">could
  </w>
  <w c5="UNC" hw="erm" pos="UNC">erm </w>
  <pause/>
  <w c5="PNP" hw="i" pos="PRON">I </w>
  <w c5="VVI" hw="mean" pos="VERB">mean
  </w>
  <w c5="NN1-VVB" hw="work"
  pos="SUBST">work </w>
  <w c5="PRP" hw="against"
  pos="PREP">against </w>
  <w c5="NP0" hw="france"
  pos="SUBST">France </w>
  <w c5="PRP" hw="in" pos="PREP">in </w>
  <w c5="DT0" hw="that" pos="ADJ">that
  </w>
  <w c5="NN1" hw="situation"
  pos="SUBST">situation </w>
  <align with="DCKLC003"/>
  <unclear/>
  <align with="DCKLC004"/>
  <c c5="PUN">?</c>
 </s>
</u>
```

The BNC licence

Originally agreed with the publishers of the copyrighted content in the corpus, the BNC licence predates modern open access licences. The licence allows distribution of the corpus by the BNC Consortium, and permits commercial use, publication of fragments of language (e.g. single sentences and concordance lines), derived wordlists, and statistical information, but does not allow re-publication or sharing of the whole corpus or whole texts. The corpus is now made available on behalf of the BNC Consortium for download for free from the Oxford Text Archive (the CLARIN-UK repository).

BNC Consortium

Chambers UCREL Longman OXFORD UNIVERSITY PRESS BRITISH LIBRARY

Oxford University Computing Services
www.oucs.ox.ac.uk

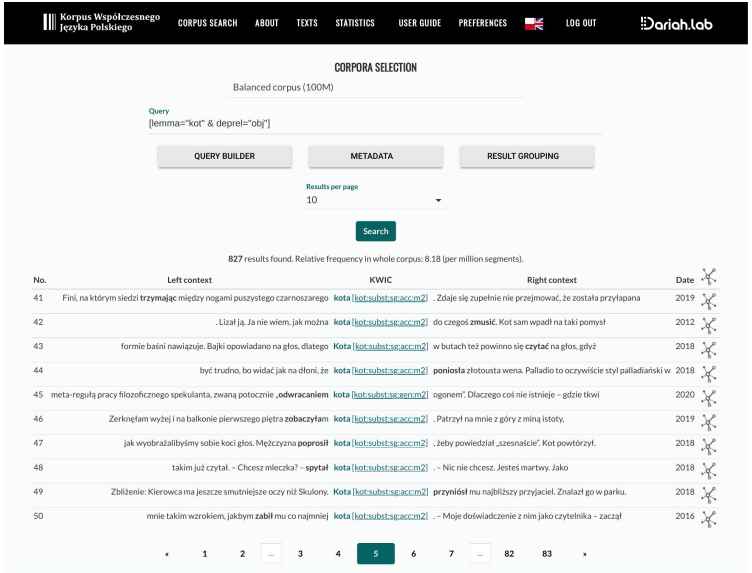
Acknowledgements

CLARIN-UK
UKRI Arts and Humanities Research Council

The Corpus of Contemporary Polish: 2011-2020 Decade and Beyond

Korpus Współczesnego Języka Polskiego

- New reference corpus of written Polish.
- Covers the 2011-2020 period.
- Wide selection of texts: fiction, non-fiction, newspapers and magazines.
- Rich annotation.
- Soon to be extended to 2020-2025 period.



https://kwjp.pl

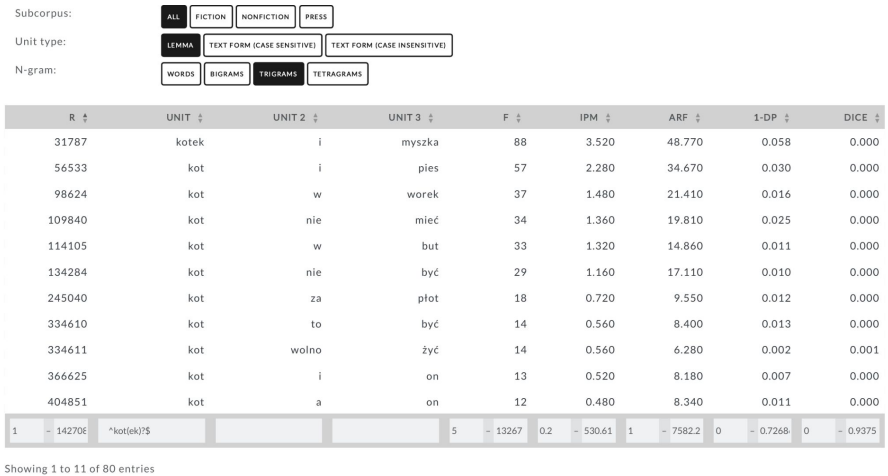
The Corpus of Contemporary Polish: 2011-2020 Decade and Beyond

Automatic annotation layers:

- Segmentation.
- Lemmata.
- Morphosyntactic tags.
- Hybrid trees (constituency-dependency).
- Named Entities.

Extended statistics:

- Frequency lists.
- Syntactic collocations.



Subcorpus: ALL | FICTION | NONFICTION | PRESS

Unit type: LEMMA | TEXT FORM (CASE SENSITIVE) | TEXT FORM (CASE INSENSITIVE)

N-gram: WORDS | BIGRAMS | TRIGRAMS | TETRAGRAMS

R	UNIT	UNIT 2	UNIT 3	F	IPM	ARF	1-DP	DICE
31787	kotek	i	myszka	88	3.520	48.770	0.058	0.000
56533	kot	i	pies	57	2.280	34.670	0.030	0.000
98624	kot	w	worek	37	1.480	21.410	0.016	0.000
109840	kot	nie	mieć	34	1.360	19.810	0.025	0.000
114105	kot	w	but	33	1.320	14.860	0.011	0.000
134284	kot	nie	być	29	1.160	17.110	0.010	0.000
245040	kot	za	płot	18	0.720	9.550	0.012	0.000
334610	kot	to	być	14	0.560	8.400	0.013	0.000
334611	kot	wolno	żyć	14	0.560	6.280	0.002	0.001
366625	kot	i	on	13	0.520	8.180	0.007	0.000
404851	kot	a	on	12	0.480	8.340	0.011	0.000
1 – 142708	^kot(ek)?$			5 – 13267	0.2 – 530.61	1 – 7582.2	0 – 0.7268	0 – 0.9375

Showing 1 to 11 of 80 entries

Building the v4 of the Croatian National Corpus (HNK)

- Previous versions of HNK published in
  - v3: 2013, 216 Mw, MSD-tagged, lemmatised
  - v2.5: 2009, 101 Mw, MSD-tagged, lemmatised
- Preparatory phase for v4 – considering issues like
  - new text sources and genres
  - coverage of different language varieties
  - more elaborated metadata
  - new structural and linguistic annotation schemata
- CorpRepo
  - a custom management system for collected corpus data developed
  - enables sustainable long-term maintenance of the collected data
  - enables publishing new versions
- Metadata: new model
  - First: metadata imported
  - Second: the actual objects harvested from respective repositories

Building the v4 of the Croatian National Corpus (HNK) 2

- Text types and IPR
  - only IPR-cleared texts: freely available for academic purposes under permissive licenses
  - collected from publicly available sources, e.g.
    - Croatian Web Archive
    - Official Journal and Central Catalogue of the Official Documents of the Republic of Croatia
    - Portal of Croatian Scientific Journals (Hrčak): 572 journals, more than 322,000 articles
    - Repositories of BA, MA/MSc, PhD theses
- Linguistic Annotation
  - HNK v4 will discontinue MulTextEast tagset and replace it with the UD tagset adding NE and syntactic annotation layers
- HNK v4
  - target size: at least 1 Gw
  - will be accessible through HR-CLARIN concordancer at https://korpusi.hr and SketchEngine
  - expected publication: end of 2026 or early 2027

Introducing the Hungarian National Corpus 3.0

Noémi Ligeti-Nagy[1], Enikő Héja[1], Ágnes Bánfi[1], Flóra Földesi[1], Bence Sárossy[1], Boglárka Skrabák[2], Tamás Váradi[1], Gábor Prószéky[1]

[1] ELTE Research Centre for Linguistics, Budapest, Hungary
[2] ELTE Faculty of Informatics, Budapest, Hungary

LREC CMLC-12 Workshop

Motivation

- Need for large-scale, curated and metadata-rich Hungarian corpus infrastructure
- MNSZ3 extends the Hungarian Gigaword Corpus (MNSZ2) from 1B to ~10B tokens
- Focus on scalability, reproducibility and sustainable digital presence
- Bridges traditional balanced corpora and modern monitor corpora

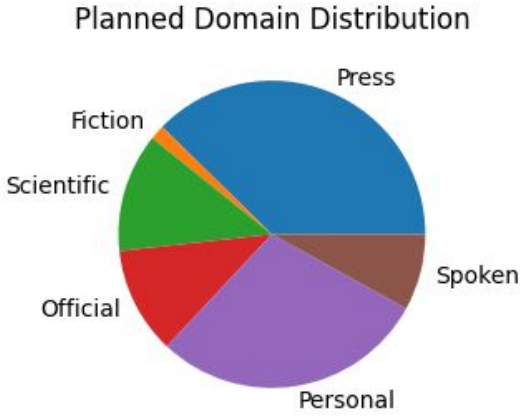
Core Objectives

- Controlled scaling while preserving balanced domain proportions
- Improved regional coverage of Hungarian outside Hungary
- Reproducible workflows for collection, cleaning and annotation
- Rich metadata and linguistically analysed content

Corpus Architecture

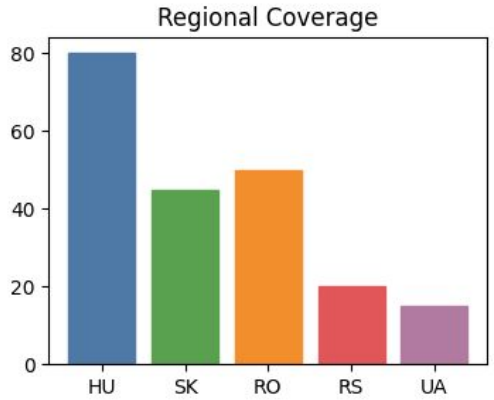
- Six domains:

  press, fiction, scientific, official, personal, spoken
- Domain-aware metadata and subcorpus queryability
- Targeted harvesting instead of unrestricted web crawling
- Deduplication, boilerplate removal and quality filtering



Scale and Regional Coverage

- Target size: approximately 10.4 billion tokens
- Expanded representation from Slovakia, Romania, Serbia and Ukraine
- Strong growth in press, official and spoken domains
- Balanced inte varieties across regions



Annotation Strategy

- Hybrid NLP workflow combining HuSpaCy and e-magyar
- HuSpaCy: tokenisation, dependency parsing, NER
- e-magyar / emMorph: morphology and lemmatisation
- Unified evaluation framework for large-scale consistency

Key Contributions

- Largest curated Hungarian corpus initiative to date
- Metadata-rich and reusable research infrastructure
- Supports corpus linguistics and NLP development alike
- Designed for sustainability, transparency and future extensibility

Takeaway

MNSZ3 establishes next-generation infrastructure for Hungarian NLP

Combines scale, balance and linguistic depth

Enables robust research on variation, genre and language technology

Thank you!

CoRoLa 2.0: corpus enrichment and a new annotation level

CoRoLa 1.0 (2017):

- 1.25B+ tokens, multi-domain, multi-genre corpus
- Text and oral

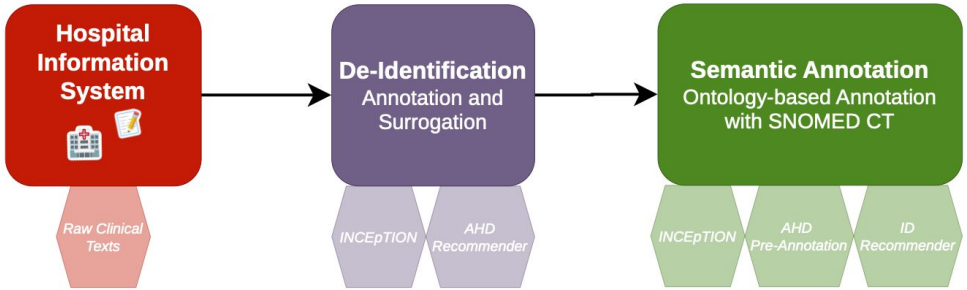
Corpus Expansion

- New data via EU & national projects
- +77M legal texts already added (LLMs4EU); more to come
- Academic & curated sources (ongoing)
- Romanian + Moldovan variety (including oral)
- Hopefully, Another 1B tokens will be added

CoRoLa 2.0: corpus enrichment and a new annotation level

- **New Annotation Layer:** Dependency parsing (RODNA pipeline; POS tagging: 97.5% | strong parsing accuracy)
- **Impact:**
    - already supported:
        - Comparative linguistic research
        - comparable corpora development
    - In ongoing projects, supports:
        - LLM development
        - AI-text and fake-news detection

The German Medical Text Corpus: Early 2026 Update



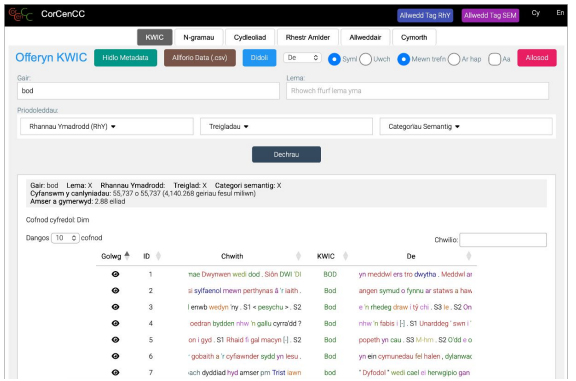
- Creating the largest German Clinical Text Corpus
  (> 18k documents, > 25M tokens)
- Six German University Hospitals & 12 methodological partners
- Tackling De-Identification & Semantic Annotation

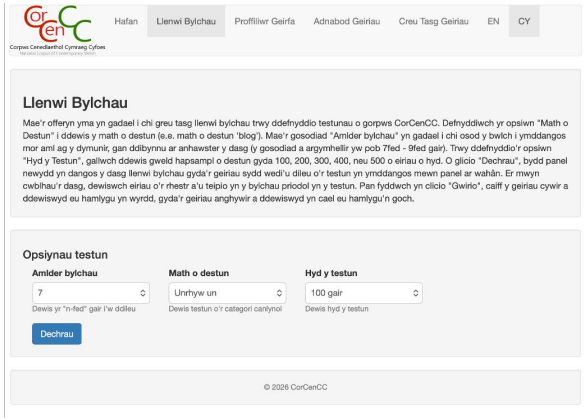
From Corpus to Community: New NLP Tools for Welsh Language Research and Learning

CorCenCC

- is the first large-scale corpus integrating spoken, written, and digital Welsh language data (11.2m words).
- includes contributions from over 2,000 speakers reflecting diverse regions and communication styles of Wales.
- enables exploration of real language use, aiding linguistic research, lexicography, translation, and language policy.

Since its release in 2020, CorCenCC has underpinned the development of a number of NLP tools and pedagogical resources for the Welsh language, providing a growing ecosystem of interoperable tools for language learning, teaching and research.





NLP and pedagogic tools built from CorCenCC

Core NLP Components

- CyTag and CySemTag provide detailed grammatical and semantic tagging for the CorCenCC.

Pedagogic Toolkit

- *Y Tiwtiadur* supports teachers and learners with corpus-based examples and interactive language tasks.

Vocabulary Development Tool

- *Yr Amliadur* offers curated, frequency-based wordlists organised by mode and part of speech.

Accessibility and Sustainability

- All tools are freely accessible via the DigiGrid platform, ensuring wide usability and long-term access.

Extending impact and democratising language technology

Geirfan Wordlist Development

- *Geirfan* is a frequency-driven wordlist tailored to A1-level Welsh learners, supporting curricula and assessments with real corpus data.

ACC Summarisation Tool

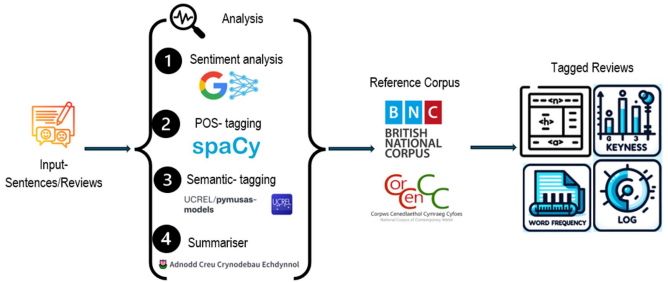
- The ACC tool generates concise summaries of long Welsh texts, enhancing education and public access to information.

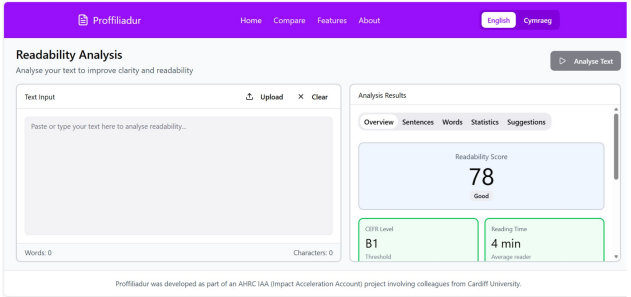
FreeTxt Data Analysis Interface

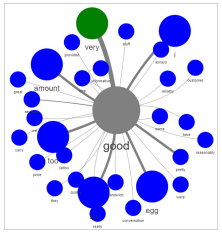
- FreeTxt combines datasets and taggers for qualitative analysis in Welsh and English through a user-friendly co-designed platform.

Proffiliadur Readability Toolkit

- Proffiliadur offers the first Welsh text-profiling toolkit, providing linguistically grounded, reproducible readability measures.







Swiss-AL: Language Data Platform for Applied Sciences

https://commons.wikimedia.org/wiki/File:Swiss_mountains,_grass_and_cows.jpg
https://commons.wikimedia.org/wiki/File:Lauterbrunnen_Bern.jpg
https://commons.wikimedia.org/wiki/File:Luzern_-_Mount_Pilatus_-_March_2019_(01).jpg

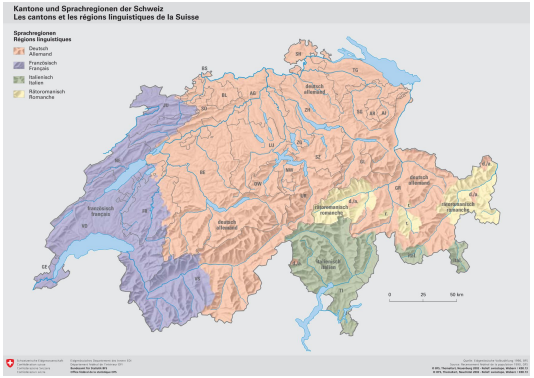
Swiss-AL: Language Data Platform for Applied Sciences

- multilingual, comparative analysis of public discourse on Switzerland
- journalistic, organizational and parliamentary corpora in all 4 national languages (DE, IT, FR, RM)
- Developed at Zurich University of Applied Sciences (ZHAW), part of CLARIN-CH
- swiss-al.zhaw.ch

Swiss-AL
LANGUAGE DATA PLATFORM
FOR APPLIED SCIENCES

zh
aw

CLARIN-CH
Common Language Resources and Technology Infrastructure



DESIGN RESPONSE 1

Make heterogeneity feel simple

01 Heterogeneous user group

scenario-based open educational resources · straightforward terminology · focus on visualisations

02 Sources and genres

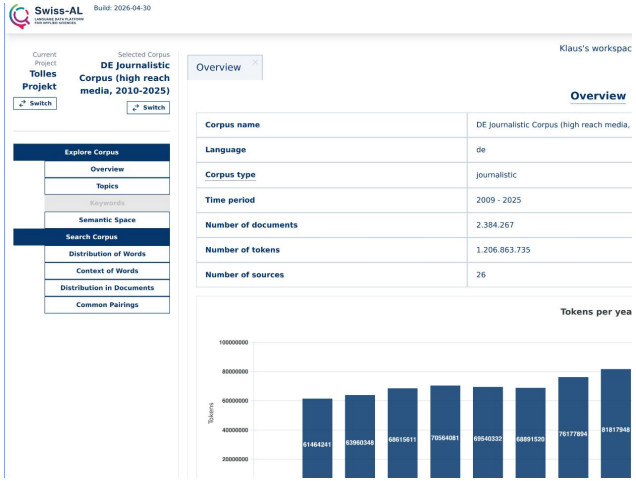
organisational, media and parliamentary discourse · metadata filtering by source, actor and time

03 Multilingual comparison

shared analysis workflows across Swiss language regions

04 Usability at scale

browser-based access modes for large corpora



Workbench interface as the user-facing response

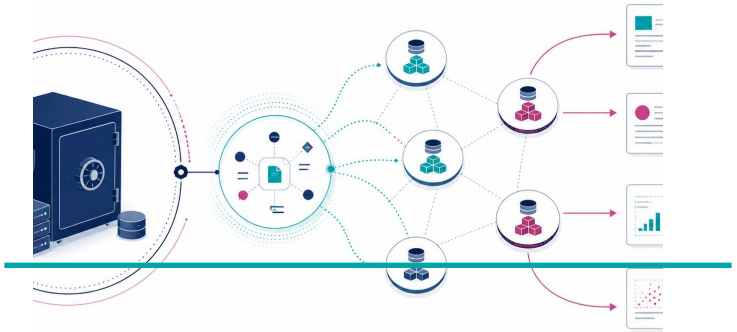
DESIGN RESPONSE 2

Protect rights while staying FAIR

For restricted corpora, “open” means discoverable, transparent and reusable metadata—not uncontrolled bulk release.

Copyright and data protection

- no bulk downloads of corpora
- short document snippets in the workbench
- external links where available
- research-only access



FAIR Open Research Data

F	**Findable**	national and European repositories
A	**Accessible**	dedicated workbench access
I	**Interoperable**	standardised export formats
R	**Reusable**	links to original data

EuReCo, KorAP and DeReKo

Updates on Ingestion and Annotation Pipelines, Backend, Interfaces, Operation, and Corpora

IDS | LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

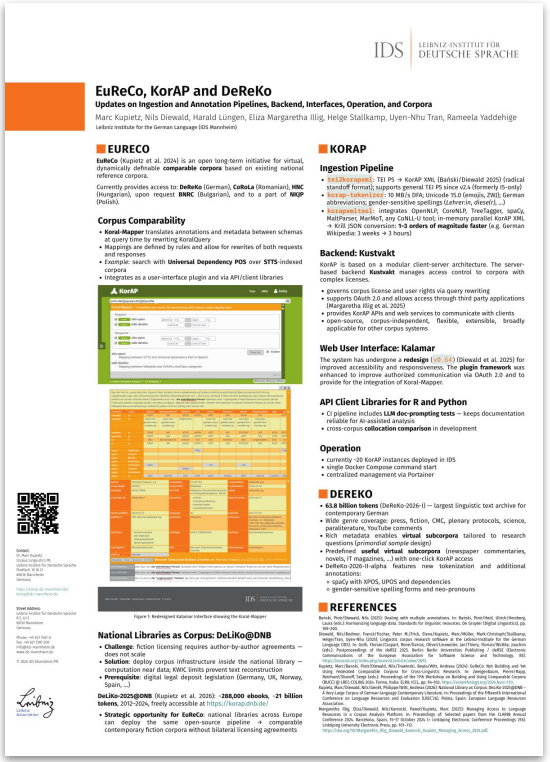
- **EuReCo**
  long-term initiative for virtual comparable corpora on top of existing national reference corpora
  - **DeReKo** (German), **CoRoLa** (Romanian), **HNC** (Hungarian), upon request **BNRC** (Bulgarian), and to a part of **NKJP** (Polish)
- **Koral-Mapper**
  annotation & metadata translation at query time
- **korapxmltool**
  fast annotation + conversion
- **Web Frontend**
  Redesigned for accessibility and responsiveness

EuReCo, KorAP and DeReKo

Updates on Ingestion and Annotation Pipelines, Backend, Interfaces, Operation, and Corpora

IDS | LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

- **DeReKo**
  - Now **63.8B tokens**
  - Newspapers, CMC, scientific articles …
  - **But:** Contemporary fiction very limited
- **DeLiKo@DNB**
  - Now **~288k ebooks, 21B tokens**
  - korap.dnb.de



IDS | LEIBNIZ-INSTITUT FÜR DEUTSCHE SPRACHE

EuReCo, KorAP and DeReKo

Updates on Ingestion and Annotation Pipelines, Backend, Interfaces, Operation, and Corpora

Marc Kupietz, Nils Diewald, Harald Lüngen, Eliza Margaretha Illig, Helge Stallkamp, Uyen-Nhu Tran, Rameela Yaddehige

Leibniz Institute for the German Language (IDS Mannheim)

EURECO

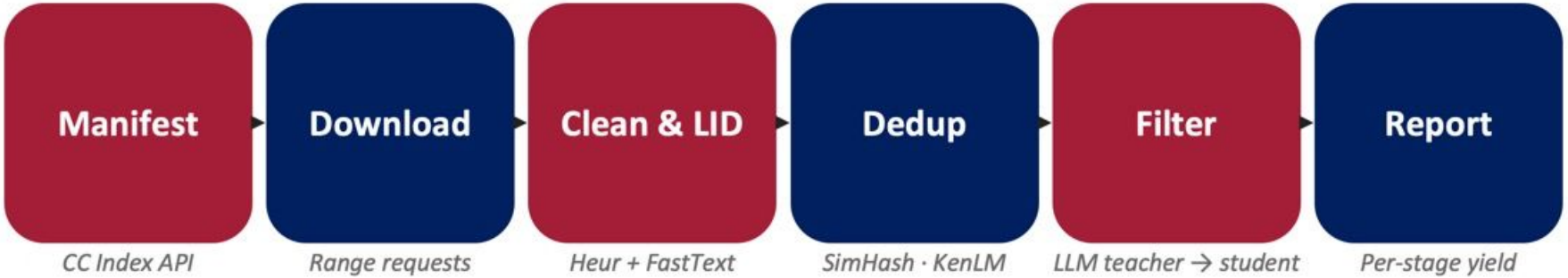
KORAP

DEREKO

REFERENCES

Merimënga: A Manifest-First Pipeline for Reproducible Albanian Web Corpus Construction

A reproducible Albanian web corpus from Common Crawl — built and re-buildable on a single workstation.



The release artefact is the recipe — manifests + scripts + ledgers — so anyone can rebuild the corpus.

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

FAU

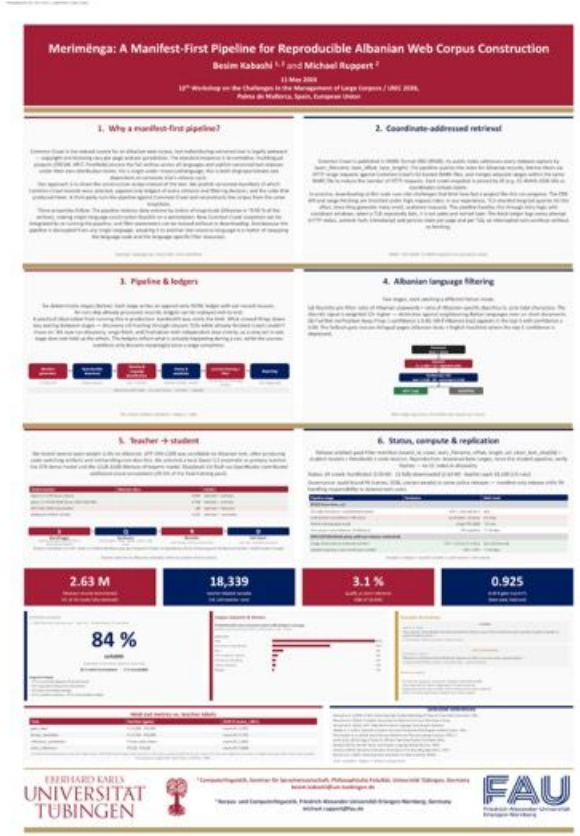
Status & next steps

Status

- Pipeline auto-cleans pages → LID → dedup → quality filter, all reproducible
- Distilled student filters: FastText for LID, XLM-R-large for quality
- Both trained from an LLM-teacher ensemble (no hand-written rules)
- Albanian demo: 2.63 M records from 24 CC snapshots

Next steps

- Re-run pipeline on newer CC snapshots, crawled after our training cutoff
- Confirm filter quality stays stable on those out-of-distribution snapshots



EBERHARD KARLS UNIVERSITÄT TÜBINGEN

FAU

The End is a Beginning

Thanks to all the presenting teams for introducing their posters!

It is now time for us to head out, grab a coffee, and drift towards Menorca Hall (3rd floor), where the posters are waiting…

Please be sure to be back in this room (room 7) *before* 11:30!

Explanation (this slide will be gone, eventually)

- Ordering of presentations matters (because time matters). Please find your presentation title below and add as many slides (**Ctrl+M**) as you need.
  - See the workshop programme for a hint on the order of presentations.
- Each group gets **150 seconds** (hard limit). These slides should therefore only contain the highlights.
- Please keep the letters large (they need to be read on the fly, there's no coming back to a slide in the discussion period – that's at the posters).
- Please rehearse: we will have to stop you after the time limit is reached, to make sure that everyone gets a chance to present.
- Do not expect this slide deck to be presented live. Since this is open to the world, and since we don't know what to expect in terms of local wi-fi, the deck is going to be downloaded *early* on May 10th and presented locally as PDF.