INNOVATION IN LANGUAGE LEARNING

International Conference

CONFERENCE PROCEEDINGS

Edited by Pixel

13th International Conference
Virtual Edition
12-13 November 2020

PIXEL

filo diritto editore
DAL 2008

CONFERENCE PROCEEDINGS

13ᵗʰ International Conference Innovation in Language Learning

Virtual Edition

Florence, Italy, 12-13 November 2020

Chapter: ICT for Language Learning

Beyond Conventional Digital Approaches in the Humanities: Corpus Linguistics for Language Related University Disciplines
Madalina Chitez, Roxana Rogobete

Cite as:

Beyond Conventional Digital Approaches in the Humanities: Corpus Linguistics for Language Related University Disciplines

Madalina Chitez[1], Roxana Rogobete[2]
West University of Timisoara, Romania[1,2]

Abstract

Diving into digital education seems to be the reality in recent times. Universities and humanities departments in particular have willingly embraced technology-rich teaching methods for a long time now [1]. Nevertheless, the general development is towards the implementation of basic digital literacy approaches rather than towards creating new discipline-adapted content that can improve the students' digital-intensive research skills [2]. This paper illustrates how the discipline of corpus linguistics can create fruitful synergies between language related disciplines in the humanities: literatures studies, both in the native and foreign languages, translation studies, gender studies, media studies, cultural studies and linguistics. The aim of the study is to demonstrate the potential of exposing students in the humanities to digital-intensive research methods, such as corpus linguistics, to enhance learning motivation and improve disciplinary competence at the same time. The first part of the study is dedicated to the presentation and exemplification of the teaching and student research methods that have been implemented with first- and second-year undergraduate students at a Romanian university who conducted a three-month internship at the Centre for Corpus Related Digital Approaches to Humanities (CODHUS). The second part includes details on the teaching scenarios and outcomes of pedagogical interventions, based on the use of corpora, in two undergraduate courses. The paper concludes with a discussion on the impact of applied corpus linguistics methods for modern teaching in the humanities as well as their integration into the larger field of digital humanities.

Keywords: applied corpus linguistics, digital humanities, innovative teaching, ICT for language learning, Romanian university students

1. Introduction

Digital approaches have gradually entered academic usage due to a multitude of society-driven factors such as the emergence of technology-supported activities in all sectors of life. It is now unconceivable to attend a lecture or finish an assignment without a variety of electronic means that either enhance communication or assist during task completion processes [1]. Policy making strategies, such as the ones promoted by the Bologna Process, have also adapted to the new reality and incorporate an array of measures that are relevant for digitalisation strategies Europe-wide. Nevertheless, the general development is towards the implementation of basic digital literacy approaches rather than towards creating new discipline-adapted content that can improve the students' digital-intensive research skills [2]. Until digital transformation becomes the norm, universities struggle with the integration of innovative computer-mediated methods and tools on a continuum ranging from reluctance to accept transformation

because of disciplinary traditions (e.g., humanities), to implementing only communication-related digitalisation solutions (most universities), up to creating synergies for a complete transformation of the research and teaching mission. In Romania, this polarisation is quite present, considering that the whole educational system is still undergoing reform away from a communist methodology (beginning of the 1990'), towards embracing European values (entering EU in 2007) and continuing with nowadays' challenge of completely having to switch to the online environment during the ongoing pandemic.

In the present study, we aim at capturing recent best-practice examples of introducing and applying digital approaches in research and research training in a rather traditional domain, humanities, at a Romanian University (i.e., West University of Timisoara/WUT, Romania). We report on the experience at the newly-founded centre CODHUS – Centre for Corpus Related Digital Approches to Humanities (https://codhus.projects.uvt.ro/), after completion of a first round of students' internship stages, and after offering two intensive undergraduate courses and one workshop for faculty members.

2. CODHUS: at the intersection of Digital Humanities and Corpus Linguistics

2.1 Digital Humanities in Romania

The results of a survey (DIGITS-Digital Humanities Survey) on Digital Humanities (DH) topics carried out last year among our contacts) show that respondents had the most diverse answers regarding the definition of the DH field, ranging from digital libraries to digital research tools [2]. But a very common answer indicated the "complementarity of technology and humanities" by "facilitating the studies and research in Humanities with the help of digital resources and tools".

Proceeding from these basic assumptions, and accounting for the relative scarcity of digital-method use in the humanities in Romania, the CODHUS research centre was founded in 2019 at WUT. CODHUS is set up as a research lab with its primary function being that of producing novel results, whereas its secondary function involves "training the next generation of scientists" [3]. A rapidly implemented measure was the open call for internships and voluntary research stages for the WUT students in humanities. This transposed CODHUS in an institution that is involved in "knowledge making" and transfer. By doing that, the lab is also building "epistemic cultures" [4], since the Romanian educational environment is very traditional (i.e., mainly instructor-centred).
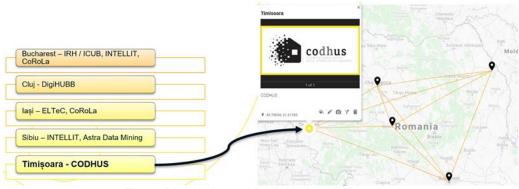

Fig. 1. Digital Humanities initiatives in Romania

Currently, in Romania, there are five universities where DH initiatives have been launched (Figure 1). Only two of them are organized as DH centres (DigiHUBB and CODHUS) whereas only CODHUS has corpus linguistics as its main focus.

2.2 Why Corpus Linguistics?

Even though nowadays the DH field extends beyond applying computational tools in linguistics or literature, language data remain probably the core area of humanities computing. Hinrichs *et al.* [5], for instance, argue that, regardless of the digital methods used, "textual resources [...] play a central role in most humanities disciplines" (p. 559).

As one of the oldest digital methods to use linguistic data, corpus linguistics "embraced the digital nearly since the inception of modern computer science" (p. 115) [6]. We therefore consider that corpus methods, either in the form of standard corpus linguistics approaches or extending towards the concept of a "corpus" as archive/collection of data, are essential in creating expertise in DH.

3. Corpus related expertise – internships and courses

3.1 A new type of internship

At WUT, internships are mandatory for most of the students (as "practice" included in the curriculum), and the educational plan states key competences that are to be developed for Philology students, such as (inter- and pluri-cultural) communication, translation competences, computer literacy skills, innovative thinking skills etc.

Thus, in November 2019, immediately after CODHUS foundation, an open call was launched and, by January, 9 interns from the 1ˢᵗ and 2ⁿᵈ year of their BA and one volunteer (3ʳᵈ year BA) were selected to start their activities at CODHUS. All of them studied English as their major or minor.

(a) The first training (February 2020) focused on the fields of DH and corpus linguistics, introducing students to a user guide in annotating and coding texts, since most of them would be involved in processing collected texts and building the ROGER bilingual corpus. This practical session was essential in accustoming trainees with linguistic data preparation and analysis.

(b) The second training (March 2020), was based on self-compiled corpora and focused on using Lancsbox [7], in case studies on expert corpora and corpus-based pedagogical interventions for the improvement of students' academic writing skills.

Trainees were taught to search frequency lists, N-Grams, keywords in contexts/concordances, GraphColls, then perform their own in-depth analysis of their findings.

(c) From March until June 2020, students were assigned several tasks under the tutors' supervision: text digitization, text processing and corpus building. Moreover, each of them chose a topic for an individual research paper, where typical corpus linguistics methods were implemented.

Discourse analysis: popular culture, media celebrities
- *Vernacular elements in song lyrics. A corpus study of Billboard hits*
- *In their own words: A study of the language of Hollywood*
- *TV language in Britain: How does it differ from "standard language"?*

Discourse analysis: media, teenagers, vlogging
- *Kallo$$ofia - Virtual friends or marketing experts? The marketing language of influencers*
- *The language of teenagers. Where does "gen" come from?*

Discourse analysis: political studies
- *The language of Romanian administration: an interview-based corpus case study*
- *Elizabeth Cady Stanton: feminist histories in documents*
- *The meaning of the word "people" in Lincoln's speeches*

Literary studies
- *Depiction of diseases in poetry: plague, cholera, COVID-19. A corpus-based study*

Urban culture studies, multi-modal approach
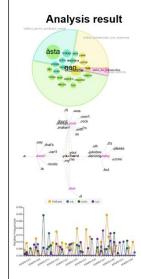- *Harlem: photographs of change*

Fig. 2. Research topics chosen by CODHUS interns and volunteers

After the internship, in their reports (online <u>here</u>), students mentioned that they perceived their internship as rewarding and helping them understand the potential of doing research using digital approaches, something that they had considered, before their experience at CODHUS, to be incompatible with the profile of the discipline.

3.2 Topics and results

As mentioned before, all the trainees were asked to carry out their own exploratory study. In order to assist students during corpus collection and analysis process, or to discuss methods and results, CODHUS tutors organized individual online meetings. An overview of the interns' research topics (Figure 2) shows that most of the studies included English language discourse analyses. Students compiled their own rather smaller (~100 texts) corpora of texts, using resources available online, several extracted from digital libraries.

Whether discussing linguistic features of Billboard's hits, those of common musical themes, or particularities in the speech of Hollywood actors in interviews, students tended to analyses phenomena related to real life or their personal interests (e.g., discourse strategies in Hollywood).

Fig. 3. Research results in internship papers (examples)

Trainees demonstrated awareness of the language transformations (e.g., examining teenagers' popular vlogs), but also well anchored in the political domain, by scrutinizing political strategies reflected in discourses of either the Romanian or the American administration. The literary topics were not neglected: one of the papers focused on differences in depicting plague, cholera and coronavirus in poetry. Another paper (cultural studies topic) designed a multi-modal analysis of the effects of urban revitalization through a collection of nine photographs depicting Harlem from the 1930s to the 2010s.

3.3 Corpus linguistics for conventional language courses

Besides internships, CODHUS offered an introductory course for both students (i.e., freshmen) and faculty members, in two sessions: in 2019 and 2020. This happened because the centre gained in popularity and more and more colleagues teaching languages manifested their interest in letting their students learn about and experiment with the new digital methods. The offer included: a module on *Digital Tools for Academic Writing* for our colleagues from different departments, and, for students, an *Introductory Course in Corpus Linguistics* (1ˢᵗ semester) and a course on *Corpus Based Literary Studies* (2ⁿᵈ semester).

Each course session ended with a student satisfaction online survey which indicated their high interest in the topic [8]. This demonstrates, once again, that the integration of new research-based digital methods in traditional language courses can be performed successfully, having, at the same time, high motivational impact.

Conclusions

After only one year of activity, CODHUS succeeded in its mission in more than one way: it became a place of encounters for students, teachers and researchers that manifested interest in the use of digital methods for the humanities disciplines, while also raising awareness, at institutional level, on the potential of language related cross-disciplinary studies. In this paper, we have exemplified the multitude of research-support

training activities that a DH centre can perform without much financial and administrative resources. Philology students performing their internships at CODHUS have reported a positive and rich learning experience:

I also found the individual research part interesting because I have never had the opportunity to do such an activity before. I believe that the programs used and the knowledge gained during the internship will be useful in the future.

(Internship Report by D. Prohap, undergraduate intern at CODHUS in 2020).

REFERENCES

[1]   Harrison, T.M.; Stephen, T. "On-line disciplines: Computer-mediated scholarship in the humanities and social sciences", *Computers and the Humanities*, vol. 26, 1992, pp. 181-193.

[2]   Chitez, M.; Rogobete, R.; Foitos, A. "Digital Humanities as an Incentive for Digitalisation Strategies in Eastern European HEIs: A Case Study of Romania", Curaj, A., Deca, L.; Pricopie, R. (eds.), *European Higher Education Area: Challenges for a New Decade*, Springer Nature, 2020 (in press).

[3]   Bangera, G., and Brownell, S. E. "Course-Based Undergraduate Research Experiences Can Make Scientific Research More Inclusive", CBE-Life Sciences Education, vol. 13, 2014, pp. 602-606.

[4]   Knorr Cetina, K. *Epistemic cultures: how the sciences make knowledge*. Cambridge/London: Harvard University Press, 1999.

[5]   Hinrichs, E.; Hinrichs, M.; Kübler, S.; Trippel, T. 2019. "Language technology for digital humanities: introduction to the special issue", *Language Resources and Evaluation*, Vol. 53, 2019, pp. 559-563.

[6]   Ebensgaard Jensen, K. "Linguistics and the digital humanities: (computational) corpus linguistics", MedieKultur. *Journal of media and communication research*, Society of Media researchers In Denmark, Vol. 57, 2014, pp. 115-134.

[7]   Brezina, V., McEnery, T., & Wattam, S. *LancsBox*, 2015.

[8]   Chitez, M.; Bercuci, L. (2019). "Data-driven learning in ESP university settings in Romania: multiple corpus consultation approaches for academic writing support". Meunier, F., Van de Vyver, J.; Bradley, L. & Thouësny, S. (eds)*. CALL and complexity – short papers from EUROCALL*. Research-publishing.net, 2019, pp. 75-81.