UNIVERSITY POLITEHNICA OF BUCHAREST

Faculty of Automated Control and Computers

PHD THESIS

Assessing Writing and Student Performance using Natural Language Processing and a Dialogical Framing

Author: Ing. Maria-Dorinela Dascălu (Sîrbu)

PhD Advisor: Prof.dr.ing. Ştefan Trăuşan-Matu

PHD COMMITTEE

President	Prof.dr.ing. Florin Pop	from	University Politehnica of Bucharest, Romania
PhD Advisor	Prof.dr.ing. Ştefan Trăuşan-Matu	from	University Politehnica of Bucharest, Romania
Member	Conf.dr.ing. Traian Rebedea	from	University Politehnica of Bucharest, Romania
Member	Prof.dr. Philippe Dessus	from	University of Grenoble Alpes, France
Member	Prof.dr. Nicolae Nistor	from	Ludwig Maximilian University of Munich, Germany

Abstract

Natural Language Processing (NLP) is a popular area of research in computer science, artificial intelligence, and linguistics domains that explores how computers can be used to understand, manipulate, and interpret text or speech in natural language. NLP includes techniques for modeling and interpreting human language, and ranges from statistical methods, machine learning, deep learning, to rule-based approaches. This thesis focuses on Natural Language Processing applications, more specifically on writing assessment, collaborative learning, and computational discourse analysis. Advanced Natural Language Processing and Machine Learning techniques are applied to analyze writing style, assess online student performance, and explore dialogism as a discourse model using language models.

The research questions of this thesis cover: (1) To what extent are textual complexity indices predictive of differences in writing style when performing multi-lingual analyses on different corpora and discourse types? (2) To what extent do features generated by the *ReaderBench* framework, including CNA applied on forum discussions, textual complexity indices, longitudinal analysis, and features derived from click-stream data, predict student performance? (3) What are the differences in students' behaviors and interactions before and during the COVID-19 pandemic in two consecutive yearly installments of an undergraduate course? (4) To what extent are semantic chains operationalized using language models predictive in both collaborative and individual settings? To respond to all these research questions, this thesis introduces a series of experimental studies.

First, this thesis analyzes the writing style in terms of linguistic features while considering Romanian media discourses on the economic crisis spanning the years between 2008 and 2018, as well as military discourses of NATO official documents. Statistical analyses were conducted to examine differences in writing styles, one-way analyses of variance to identify the most predictive textual complexity indices, followed by a stepwise discriminant analysis.

Second, this thesis evaluates students' online activity by analyzing their behaviors, modeling their interaction, and predicting students' grades based on their online participation; these case studies address the second and third research questions. Global features (e.g., participation indices, initiation indices), time series analysis and click-stream data are used to predict student grades and success. Various sociograms were generated to model trends in students' participation and to visualize the interaction patterns between participants, together with weekly snapshots to follow how the community evolves from one week to the next. Linear models indicated that math success was related to days spent on the forum and by students who more regularly posted in the online class forum. Students' behaviors and interactions are compared before and during COVID-19 using two consecutive yearly instances of an undergraduate course. The COVID-19 outbreak generated an off-balance, a drastic increase in participation, followed by a decrease towards the end of the semester, compared to the previous academic year when lower fluctuations in participation were observed.

Third, this thesis introduces a method grounded in dialogism which identifies semantic links using only the attention scores computed using BERT language model; this model supports the last research question. Although the differences in performance are not substantial, the model has important advantages compared to other methods: the chains computed by the BERT are context dependent, the model generalizes across multiple relations, and it detects other relations that share similar attentions scores, even though the prediction models were trained on simple rules. The model was applied on two tasks: automated essay scoring and assessing student involvement in chat conversations. The results showed that the ratio of chains, denoting the coverage of semantic chains used by a participant in relation to the entire conversation, is by far the most predictive feature for participation. Moreover, continuations, reflective of links between two different participants, coupled with counts and ratios of chains covered by the participant are the best predictors for collaboration.

Table of Contents

List of Figures

1 Introduction

1.1 Overview

Natural Language Processing (NLP) (Chowdhury, 2003; Jurafsky & Martin, 2008) and Machine Learning (ML) (Jordan & Mitchell, 2015; Mohri, Rostamizadeh, & Talwalkar, 2018) are gaining an increased attention due to the increasing volume of data that exists and needs to be processed. NLP explores how computers can be used to understand, manipulate, and interpret human language. NLP tasks and applications vary greatly, for example: question and answering, text and question generation, sentiment analysis, fake news detection, dialog understanding, text summarization, reading comprehension, keyword extraction, topic modeling. This thesis presents a wide range of experiments based on NLP techniques as a field of activity, with focus on modeling writing style, assessing online student performance in online learning environments, and exploring dialogism using language models.

Writing is a central skill needed for learning that is tightly linked to text comprehension. Good writing skills are gained through practice and are characterized by clear and organized language, accurate grammar usage, strong text cohesion, and sophisticated wording.

Online collaborative learning environments open new research opportunities, for example, the analysis of learning outcomes, the identification of learning patterns, the prediction of students' behaviors, and the modeling and visualization of social relations and trends among students. Various online learning platforms are available for free to assist students in the learning process. The best known are Moodle (Moodle, 2021) and Massive Open Online Course (MOOC; McAuley, Stewart, Siemens, & Cormier, 2010). Moodle is an online educational platform designed to enhance both student and teacher activity and can be effectively employed to support collaborative learning. Moodle is often used to make inquiries on student homework, exams, to request clarifications, and to make announcements. MOOCs have become an important platform for teaching and learning because of their ability to deliver educational accessibility across time and distance. Assessing student success and involvement in online learning environments is a difficult and time-consuming task for teachers. One data tool that has been proven effective in exploring student success in on-line courses has been Cohesion Network Analysis (CNA; M. Dascalu, Trausan-Matu, McNamara, & Dessus, 2015), which offers the ability to analyze discourse structure in collaborative learning environments and facilitate the identification of learner interaction patterns. These patterns can be used to predict students' behaviors such as dropout rates and performance.

Dialogism is a philosophical theory centered on the idea that life involves a dialogue among multiple voices in a continuous exchange and interaction. Considering human language, different ideas or points of view take the form of voices, which spread throughout any discourse and influence it. From a computational point of view, voices can be operationalized as semantic chains that contain related words.

1.2 Goals and Interests

Based on the previous context and the multiple areas of research it provides, this thesis focuses three main objectives, out of which four research questions arise. As such, the objectives of this thesis are threefold, and each objective is detailed in one dedicated chapter from the empirical studies part of this thesis:

1. Modeling and investigating writing style in various texts using textual complexity indices;
2. Evaluating the online activity of students in online learning environments based on their posts (e.g., Moodle courses, MOOCs) by analyzing students' behaviors, modeling their interaction, and predicting students' grades based on their participation;
3. Introducing a novel method of identifying semantic chains using state-of-the-art language models to evaluate students' involvement, based on their writing in online chat conversations.

Based on these three main objectives, this thesis answers the following research questions:

- **RQ1**: To what extent are textual complexity indices predictive of differences in writing style when performing multi-lingual analyses on different corpora and discourse types (i.e., NATO official documents and Romanian media)?
- **RQ2**: To what extent do features generated by the *ReaderBench* framework, including CNA applied on forum discussions, textual complexity indices, longitudinal analysis, and features derived from click-stream data, predict student performance?
- **RQ3**: What are the differences in students' behaviors and interactions before and during the COVID-19 pandemic in two consecutive yearly installments of an undergraduate course?
- **RQ4**: To what extent are semantic chains operationalized using language models predictive in both collaborative (i.e., evaluating collaboration between chat participants) and individual settings (i.e., automated essay scoring)?

The first research question is directly mapped to the first objective, the second and third research questions are linked to the second objective, whereas the last research question is derived from the third objective.

1.3 Thesis Outline

The thesis is structured into three main parts, *Theoretical Aspects*, *Empirical Studies*, and *Discussions* and *Conclusions*. The chapters from the *Theoretical Aspects* part support the chapters from the *Empirical Studies* part (see Figure 1). Each subchapter in *Empirical Studies* has a corresponding chapter with theoretical aspects, which includes state-of-the-art models, methods, and systems.

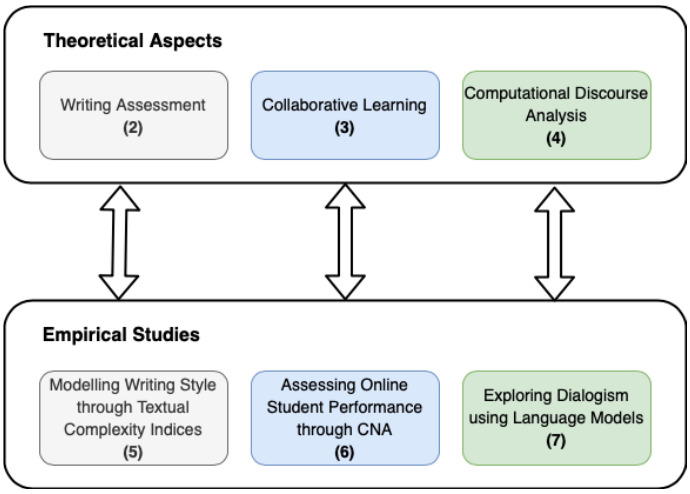


Figure 1. Thesis structure.

The thesis starts with theoretical aspects and includes state-of-the art models, methods, and systems, which are relevant for empirical studies. This part begins with writing assessment concepts (Chapter 2), defining writing as a critical task, and describing various Automated Writing Evaluation (AWE) systems. Chapter 2 continues with the definition of dialogism and the polyphonic model, and ends with aspects related to stylometry, including different state-of-the art systems for computing textual complexity indices. Chapter 3 introduces collaborative learning and focuses on online learning environments which stimulate collaboration (e.g., learning management systems, course management systems, MOOCs). This chapter introduces dialogism as a framework for Computer Supportive Collaborative Learning (CSCL) environments, end ends with Social Network Analysis (SNA) and the most common metrics to compute participation, collaboration, and importance in the community. Chapter 4 presents theoretical aspects related to computational discourse analysis, and includes general aspects about Natural Language Processing, with an emphasis on spaCy, state-of-the-art semantic models and language models, and Cohesion Network Analysis.

The thesis continues with empirical studies and define various case studies which are based on the theoretical concepts described in the first part, grouped according to their specificity. Each chapter from this section includes the description of the experiment, the corpus used, the method approached, results and discussions. Thus, Chapter 5 describes two case studies that focus on modelling writing style through textual complexity indices. Chapter 6 introduces two case studies centered on assessing online student performance through Cohesion Network Analysis, while Chapter 7 presents two case studies that explore dialogism using language models.

Next, the *Discussions* part describes advantages of our approaches, faced problems, and provided solutions, alongside with educational implications. The *Conclusions* chapter presents personal contributions and directions for future research.

Part 1 – Theoretical Aspects

2 Writing Assessment

Writing is a central learning activity that requires practice and experience (Light, 2004; McNamara, Crossley, & McCarthy, 2010; Witte & Faigley, 1981). Writing quality is an indicator of performance. A good writing has focus, consistency, coherence, correctness and development (College, 2021). Moreover, it has one clear main idea, and each paragraph has a single main sentence or a single main perspective. The focus should be on one main idea, which spreads throughout the writing. A good writing is consistent, each paragraph is related to the main idea, but have its own perspective and does not share with another paragraph. Coherence is a very important quality, as the ideas or events should be logically organized. Good writing ensures a cohesive flow of information and makes sense to the reader. Writing quality is also indicated by correctness, because a right writing is written correctly, without grammatical mistakes, the words are used correctly, the sentences are complete and make sense. The main idea is supported and expanded by each paragraph. Ideas are described in detail, explained, and supported by examples.

Good writing also requires creativity, while the author's personality is found in the way he expresses ideas and in the combination of words. Each person has his own way of expressing himself, and a more or less complex vocabulary. A good writer has clarity and discipline in his expressions.

Good writing also requires that the text is easily understood by the targeted audience. Readability (Dale & Chall, 1949) decides how difficult or easy is to understand a text. There are various elements that contribute to the readability of a text: words, length of sentences, sentence's structure, the average syllables per word (DuBay, 2004). These combined elements can make a text hard to understand. A high readability means that the reader can clearly understand the ideas of the text while accounting for the complexity of concepts, can easily process all the information provided within the text and reduces misunderstandings.

2.1 Writing as a Critical Task

Writing is a central skill needed for learning that is tightly linked to text comprehension (Moravcsik & Kintsch, 1993). Good writing skills are gained through practice and are characterized by clear and organized language, accurate grammar usage, strong text cohesion, and sophisticated wording.

Providing constructive feedback can help learners improve their writing; however, providing feedback is a time-consuming and complex process.

Many students are interested in improving writing skills because writing quality is an important aspect in defining intellectual capabilities at almost all academic levels. Providing personalized feedback to students about their writing ability is a fundamental component in the learning process. However, many teachers do not have the time to provide feedback in an iterative manner that best supports student learning. As a result, many computer-based systems have been developed to support students in the writing process. These systems are generally referred as Automated Writing Evaluation systems (AWE). Various systems capable to provide writing strategy training have been developed in order to help students in the writing process (McNamara, Crossley, & Roscoe, 2013).

According to Roscoe, Varner, Crossley, and McNamara (2013), there are two categories of textual evaluation systems that can provide feedback to students: Automated Essay Scoring (AES) systems and Automated Writing Evaluation (AWE) systems. AES systems (Attali & Burstein, 2004) provide an overall, summative score on an essay and their performance can be evaluated in terms of the correlation between human and automated scores. AWE systems (Warschauer & Ware, 2006) are built on top of AES systems with the aim to provide constructive feedback that goes beyond holistic scoring (i.e., formative in addition to summative scoring). AWE systems are more useful for student development since the received feedback can help them in terms of awareness of potential issues, as well as methods of improvement. Both AES and AWE systems can rely on textual complexity indices which provide indications regarding a text's quality or traits of writing style, and can measure different aspects ranging from readability, surface-level numerical metrics (such as paragraph counts, syllable counts, word entropy etc.), to text coherence and cohesion (Crossley, Kyle, & McNamara, 2015; McNamara, Crossley, Roscoe, Allen, & Dai, 2015).

Many AWE systems have been developed for the writing classroom. Part of these systems are freely available (in most cases, those developed by researchers), while systems developed by companies usually require a small to medium fee.

2.2 Dialogism and the Polyphonic model of Discourse

Dialogism is a philosophical theory introduced by Mikhail Bakhtin (Bakhtin, 1981, 1984), centered on the idea that everything, even life, is dialogic, a continual exchange and interaction between voices: "Life by its very nature is dialogic ... when dialogue ends, everything ends" (Bakhtin, 1984, pp. 293,252). Trausan-Matu, Stahl, and Sarmiento (2007) extended the concept of voice for analyzing discourse, in general, and collaborative learning, in particular. They consider voices to be generalized representations of different points of view or ideas, which spread throughout the discourse, and influence it. There are several voices in a dialogue that support communication purposes and interactions, inter-animating through divergences and convergences of points of view, while eventually offering harmony to the discourse, in a way similar to the polyphonic music (Trausan-Matu, 2020). Going broader, dialogism can be perceived as a theory of a meaningful world (Linell, 2009), which consists of meaningful ideas, thoughts, actions, and processes.

Utterances, either in speech or text, are multivocal, everything we say or write is filled with the words and perspectives of others, and vice versa (Bakhtin, 1986). In this sense, the concept of dialogue acquires other connotations, and can be extended to include a wider range of language activities. Thus, dialogism and inter-animation of voices are encountered in all contexts, from conversations in which multiple participants discuss on various topics, to texts in general in which different authorial positions appear – Dostoyevsky's prose perceived as being "a plurality of independent and unmerged voices and consciousnesses" (Bakhtin, 1984) – and even internal dialogues between inner voices that debate ideas and contradict each other (Marková, Linell, Grossen, & Salazar Orvig, 2007). Multivocality leads to polyphony, a central concept in dialogue, and a central point of the polyphonic model introduced by Trausan et al. (Trausan-Matu, 2020; Trausan-Matu & Rebedea, 2009; Trausan-Matu, Stahl, & Sarmiento, 2007).

Voices were previously operationalized by M. Dascalu, Trausan-Matu, and Dessus (2013) as semantic chains that were obtained by combining lexical chains, i.e., sequences of repeated or related words, including synonyms or hypernyms (Mukherjee, Leroy, & Kauchak, 2018). The way in which ideas propagate throughout the text, the type of used words, the manner of expression, are all characteristics of the writing style and define types of texts. To emphasize the characteristics of the writing style, the next section describes the stylometry and systems that help to calculate various textual complexity indices.

2.3 Stylometry

Stylometry is the study of literary style that defines the way the author uses words and the way he organizes sentences to express his ideas (Holmes, 1998). We like to express ourselves in a certain manner, varying from simple to complex words and phrases, from short sentences to long paragraphs of text with long sentences. Even if the rich vocabulary may suggest a higher literal quality, it is not always true. No one stopped Ernest Hemingway from winning the Nobel Prize in 1954 for literature, even though he used a surprisingly small number of words (Rice, 2018).

Functional words are words that do not contain textual information or have a low lexical meaning., and they are used to express grammatical relationships between other words. This type of word specifies the attitude or state of speech. Examples of functional words are conjunctions, prepositions, grammatical articles, auxiliary verbs. These words are also called stop words because they do not contain textual information. Words that are not stop words are called content words, these are words that contain textual information and include nouns, adjectives, verbs and adverbs. Functional words have proven to be quite powerful in the way they are used by authors. Stamatatos (2009) conducted a study on automatic authoring approaches. In his study, Stamatatos (2009) points out that functional words "used in a largely unconscious manner by the authors, and they are topic-independent." (Stamatatos, 2009). Thus, the unconscious use of functional words is an advantage in the analysis of stylometry because it varies less than the vocabulary of an author.

There are various systems that compute indices for textual complexity, at different processing levels: surface, lexical, syntax, semantic, discourse structure and word. *Coh-Metrix* (McNamara, Graesser, & McCarthy, 2014) is one of the first and best known systems which computes various indices at all levels. *TAALES* (Kyle & Crossley, 2015; Kyle, Crossley, & Berger, 2018) measures over 400 indices for lexical sophistication. *TAALED* (Kyle, Crossley, & Jarvis, 2020) computes a wide variety of lexical diversity indices. *TAASSC* (Kyle, 2016) measures a number of indices related to syntactic sophistication and complexity. *TAACO* (Crossley, Kyle, & Dascalu, 2019; Crossley, Kyle, & McNamara, 2016) computes 150 indices of both local and global cohesion. *TAALES*, *TAALED*, *TAASSC* and *TAACO* offers a wide range of indexes by category. *ReaderBench* (M. Dascalu, 2014; M. Dascalu, Crossley, McNamara, Dessus, & Trausan-Matu, 2018; M. Dascalu, Dessus, Bianco, Trausan-Matu, & Nardy, 2014) is an open-source advanced Natural Language Processing framework grounded in Cohesion Network Analysis (CNA) and computes various indices at all levels. *ReaderBench* is developed in our research group, tries to cover the most common features which are then integrated into subsequent processing pipelines.

3 Collaborative Learning

Collaborative learning is a social process of building knowledge through collaboration (Stahl, 2006). According to Stahl (2006), several phases are in a continuous cycle and contribute to the construction of both personal and social knowledge.

Computer Supported Collaborative Learning (CSCL) is an emerging paradigm of educational technology (Koschmann, 1996). The development of technology has shaped our conceptions of human knowledge and learning (Bolter, 1984). In CSCL, technology meets educational sciences, psychology and philosophy. According to Lipponen (2002), the first CSCL workshop was held in 1990, and the first CSCL International Conference was held in 1995 in Bloomington, Indiana. CSCL assumes that collaborative learning is supported by technology to improve interaction between participants. Technology facilitates the exchange of ideas and opinions and help the rapid distribution of knowledge.

CSCL has become increasingly used in educational contexts due to its synergic effects among peers. Learners share their ideas and opinions, learn from each other, while having access to a wide range of materials. Chats and forums, the most commonly used CSCL environments, offer learners the opportunity to work together to solve problems and ask for help when encountering issues – more generally, students collaboratively build knowledge and share it among all participants (Stahl, 2006). Collective and individual learning processes intertwine one with another to create collaborative knowledge, which is spread among all participants (Cress, 2013).

In CSCL environments, technology empowers communication and collaboration throughout the learning process. However, from tutor's perspective, analyzing the resulting conversations is a time-consuming task due to their increased volume. Therefore, automated tools that analyze conversations and evaluate collaboration between participants have become a necessity. Moreover, collaboration and creativity among peers can be stimulated using automated processes (Stamati, Dascalu, & Trausan-Matu, 2015). In addition, repeatable tasks, which do not require a human response, can be automatically performed by these systems while using conversational agents (Gabriel, Hahne, Zimmermann, & Lenk, 2021).

3.1 Dialogism as a Framework for CSCL

Dialogism is considered the theoretical framing or paradigm of CSCL (Koschmann, 1999; Trausan-Matu, Stahl, & Sarmiento, 2007), a central component to understanding how information is propagated through conversations and the flow of information in discussions (M. Dascalu, Trausan-

Matu, McNamara, & Dessus, 2015). Voices (points of view) take the form of concepts or events that are propagated throughout the conversation by participants who share convergent or divergent perspectives (Trausan-Matu, 2010). Multivocality is centered on the dialogue between multiple voices and meanings, while polyphony focuses on the inter-animation between voices and the relationships that appear by overlapping them. In the polyphonic model, voices are not associated with just one individual, they can become emergent topics or themes from the discourse. Thus, a multitude of voices inter-animate with each other, each with its own identity and musicality, but forming together a comprehensible entity (Trausan-Matu & Rebedea, 2009).

Vocal inter-animation and polyphony reflect the essence of a good CSCL conversational session, in which multiple perspectives and points of view interact, striving towards unity or diversity. Explicit and implicit references within the conversations introduce longitudinal dimensions of dialog (along time) by chronologically following the events within the discussion. The interactions between members are reflected in their voices; thus, polyphony can be perceived as an indicator of collaboration (M. Dascalu, Trausan-Matu, McNamara, & Dessus, 2015). Moreover, the manner in which voices evolve over time, their influence on other participants, their inter-animation, all provide an in-depth analysis of speech. The exchange of ideas, points of view and topics of interest, divergences and convergences reflect a transversal dimension (across time), which captures movement when members of a conversation shift to discuss other topics (Trausan-Matu, Stahl, & Sarmiento, 2007). At any time, a new topic may appear in a conversation, and several voices are present simultaneously. Therefore, a discussion represents the inter-animation of several voices, be them harmonic or dissonant. The evolution of voices throughout a conversation and their influence on other participants provide valuable insights into collaboration.

3.2 Online Learning Environments to Support Collaboration

Online Learning Environments (OLEs) are increasingly used by people around the world because they facilitate quick access to resources and information, allow users to share opinions and ideas, and even engage in open debates. Learners share their experiences and opinions and search for answers in online environments, while tutors and instructors share their knowledge and expertise. Moreover, in addition to the facilities brought to learners and instructors, OLEs have opened up new research areas, such as modeling members' participation, analyzing interactions, and identifying particular interaction patterns within the community (Moore, Dickson-Deane, & Galyen, 2011; Tu, 2002; Weidlich & Bastiaens, 2019).

Online learning is constantly evolving and has changed considerably since its first appearance. In 1989, the University of Phoenix, one of the pioneers in online education, offered the first online program which became the school's main focus (Britannica, 2021). Ten years later, the first entirely web-based university – Jones University – became accredited and, only one year later, the University of Texas provided a number of online classes on a website that included quizzes, surveys, grades, and calendars. Thus, a new industry emerged – online education technology. Dave Cormier from the University of Prince Edward Island coined the term Massive Open Online Course (MOOC) in 2008. Various MOOC environments appeared in subsequent years, including Coursera, edX, and Udacity (Achieve Virtual, n.d.). Online learning improved distance learning and began to act as a substitute for face-to-face classes. As technology continues to evolve at a rapid pace, institutions striving to provide high quality education are required to use the best alternatives (Hiltz & Turoff, 2005).

Various types of systems have been developed to support online learning. Learning Management Systems (LMSs) (Watson & Watson, 2007) offer a complex learning environment, which is capable to manage all elements of the learning process. Course Management Systems (CMSs) (Simonson, 2007) has more limited functionality than LMSs and provide a learning environment which encourage collaboration between students. While LMSs are focused on learners and organizations, CMSs are focused on tutors and students. Massive Open Online Courses (MOOCs) are online courses available to anyone. MOOCs are free and offer distance learning opportunities for people around the world.

3.3 Social Network Analysis to Model Collaboration

Social Network Analysis is a fruitful approach to analyze social structures (SNA; Scott, 1988). Social Network Analysis, field of data analytics, uses graph theory and networks to represent and understand social structures. SNA helps in understanding the behavior of a network, how nodes are connected, what are the relations between the nodes and what are the most important nodes. A social network is a set of entities and relations between them (Knoke & Yang, 2019) and "provides a powerful model for social structure" (Scott, 1988).

Various measures and metrics can be used in the analysis of social networks, at entity (node) or network level. Centrality measures provide information about the importance of nodes and edges within a social network: degree centrality, betweenness centrality, closeness centrality, PageRank centrality, eigen centrality. A high score of centrality indicates influence, power in the social network. At network level, measures include density, network size.

Degree centrality of a node indicates the number of nodes which are directly linked to the current node (Das, Samanta, & Pal, 2018). Degree centrality is the simplest measure of a node's connectivity and identifies the nodes with the most connections to other nodes in the network. Thus, the measure of degree centrality identifies which is the most popular node in a network and which is the least popular. Nodes with a high degree centrality have the best connections and it can be powerful within the network. Degree centrality defines the number of edges for a specific node. If the network is directed, then two measures of degree centrality exist: *indegree* and *outdegree*. Indegree is the number of edges which exit the current node to other nodes in the network, while the outdegree is the number of edges which enters in the current node.

Betweenness centrality of a node *x* represents the number of shortest paths between a node source *s* and a node destination *d*, passing through node *x*. It was first introduced by Shaw (1954) for human networks and then by Freeman (1977) for social networks. Most often, nodes with a high betweenness centrality score act as "bridge" between the other nodes and create short paths within the network. Such nodes play an important role within the network as control the strongest flow of information between nodes, and its elimination causes a rather large disruption (Arif, 2015).

Closeness centrality measure represent the average distance between a node and every node in the network (Hansen, Shneiderman, Smith, & Himelboim, 2011). This measure checks how strong connected is a node with every node in the network and finds all nodes which are closest to other nodes. In order to do this, the closeness centrality measure find all the shortest paths within the network. After that, a score for each node is computed using its shortest paths.

Eigen centrality measure finds the nodes with the greatest influence within a network. Eigen centrality it's an extension of degree centrality, it starts by measuring the degree centrality for the current node, after that it measures the degree centrality of all nodes which the current node connects to, and so on, until the end of the network. The algorithm is using the power iteration method (Cambridge Intelligence, 2021).

The *Page Rank* algorithm (Brin & Page, 1998; Page, Brin, Motwani, & Winograd, 1999) computes the importance or influence of each node, based on the incoming relationships and the importance of the equivalent source nodes. Using the Page Rank algorithm, you can find out who has a wide influence at the network level or who is important within the network at the macro level. Even if Page Rank is a variant of Eigen Centrality that was designed to rank web content, using the hyperlinks between pages, it can be used for any type of network. The main difference between Eigen Centrality and Page Rank is that the Page Rank take into account the direction of the edge.

Network size represents the number of nodes within the network, it doesn't include the number of edges. *Network Density* is computed as the ratio between the number of existing edges and the total number of possible edges.

4 Computational Discourse Analysis

Natural Language Processing (Jurafsky & Martin, 2008) is a very popular area of research that explores how computers can be used to understand, manipulate and interpret text or speech in natural language (Chowdhury, 2003). Identifying how human beings understand and use written or spoken natural language underlines the development of special tools and techniques for computer systems to understand natural language. NLP is based on several disciplines, including computer science and computational linguistics. By combining computational linguistics with machine learning, statistical methods and deep learning models, computers are able to process human language, to understand it, in order to extract various features used in different NLP tasks.

Human language is very complex and diverse. Each language has its own rules, both grammatical and syntactic. Humans express themselves in many ways, very differently, both written and verbal. In written language, most of the time humans abbreviate words, make grammatical or syntactic mistakes, omit punctuation marks. In spoken language, we sometimes borrow terms from other languages, use regionalisms or different accents. Thus, in addition to modeling human language, syntactic and semantic understanding are very important for NLP tasks.

NLP tasks include speech recognition, sentiment analysis, part-of-speech tagging, word sense disambiguation, named entity recognition, co-reference resolution, natural language generation, natural language understanding. NLP applications covers a variety of areas: machine translation, chatbots, virtual assistants, spam detection, summarization, sentiment analysis, question and answering.

4.1 Natural Language Processing Pipeline

Natural Language Processing pipeline is a chain of steps that feed into each other to solve a particular problem. The common NLP pipeline consists of three steps: text processing, feature extraction and modeling. Each step returns a result which is used within the next step. The text processing step deals with text cleaning, tokenization, lemmatization, part-of speech tagging. Feature extraction gets appropriate feature representations for the desired NLP task and model used. Modeling deals with designing a model, by adapting the parameters to the data used in training and using it for predictions.

Various libraries were developed for supporting NLP tasks: Spark NLP (John Snow Labs Inc, 2021), spaCy (Explosion, 2016-2021), Allen NLP (The Allen Institute for Artificial Intelligence, 2021), Stanford Core NLP (Stanford NLP Group, 2021), Gensim (LGPLv2.1, 2009-2021), Rasa (Rasa Technologies Inc, 2021).

SpaCy (Explosion, 2016-2021) is a Python open-source library for advanced Natural Language Processing. SpaCy can be used to process or pre-process texts in order to build systems for information extraction, summarization, natural language understanding, question and answering. SpaCy offers multi-language support, and its pipelines are trained for each language. SpaCy has various features referring to linguistic notions or functionalities for general machine learning. SpaCy features include *Tokenization*, *Part-of-speech (POS) Tagging*, *Dependency Parsing*, *Lemmatization*, *Sentence Boundary Detection*, *Named Entity Recognition*, *Entity Linking, Similarity*, *Text Classification*, *Rule-based Matching*, *Training* and *Serialization*.

4.2 Semantic and Language Models

Word Representations using Semantic Models

Word representations has an important role in NLP and refers to representing a word using a vector (Liu, Lin, & Sun, 2020). Words represent the smallest significant unit used by people in speech or writing. Sentences and paragraphs contain groups of words arranged so that the resulting text makes sense and is easy to understand. In order to understand such a text, people need to understand correctly every word it contains and its meaning. However, for NLP tasks, each word needs to have a clear representation in order to help the models understand as well as possible the words meanings.

Word embeddings encode similar words with similar encoding and provides a distributed representation of a text in an n-dimensional space. An embedding is a dense vector, whose values are floating points and represent weights learned by a model. For a small data set, word embeddings are usually 8-dimensional, but they can be up to 1024-dimensional, for very large data sets. Using a larger dimension can catch deep relationships between words, but it takes a lot of data to train. Word embeddings has been used in various NLP tasks, such as: sentiment analysis, document classification, recommendation systems, text summarization.

Latent Semantic Analysis (LSA; Dumais, 2004; Landauer & Dumais, 2008; Landauer & Dumais, 1997) is a statistical and mathematical method for extracting meaning of words from texts in order to infer the concepts. LSA uses Bag of Words model (BoW) which represents a text using the occurrences of each word it contains.

Latent Dirichlet Allocation (LDA; Blei, Ng, & Jordan, 2003) is a probabilistic model for topic modeling. Within a document or collection of documents, there may be several topics. Topic modeling helps to automatically understand and organize documents collections. At the same time, modeling topics is very useful for summarizing documents content or searching for information in documents collections.

Word2vec (Mikolov, Chen, Corrado, & Dean, 2013) is a well-known method for word embeddings. Word2vec takes as input a corpus of texts and returns a set of vectors that means the vector representation of the words in the corpus. Vectors are chosen using the cosine similarity function. Word2vec is based on a neural network architecture, which consists of two models: Continuous Bag-of-Words model (CBOW) and Continuous skip-gram model (SKIP–GRAM).

Contextualized Representations using Transformer-based Language Models

Transformer (Vaswani et al., 2017) is a model of neural network architecture, which is based on a mechanism of attention, more-specific self-attention. The attention mechanism is used to extract the dependencies between input and output by looking at the entire input, not only at the latest segment. For example, if we have the following input "The cat saw a mouse. It was in the mood to play, so it tried to catch it, but it got into the hole in time.", the attention mechanism can detect who each "it" used refers to. According to Vaswani et al. (2017) "Transformer is the first transduction model relying entirely on self-attention to compute representations of its input and output without using sequence–aligned Recurrent Neural Networks (RNNs) or convolution".

Bidirectional Encoder Representations from Transformers (BERT; Devlin, Chang, Lee, & Toutanova, 2018) is a language representation model, which brought the latest state-of-the-art results in various NLP tasks, such as text classification, question and answering, text summarization, sentiment classification, sentence prediction, natural language inference, word sense disambiguation. BERT was designed to pre-train deep bidirectional representations from unlabeled texts, by using masked language models. BERT is based exclusively on the attention mechanism. Words change their meaning as the sentence or paragraph develops. Focus words become more ambiguous if the sentence or paragraph contains more words, and their meaning increases. BERT explains the effect that other words have on the focus word. Being a bidirectional model, during the training phase BERT learns information from both the left and the right side of the context of a certain symbol.

4.3 Cohesion Network Analysis (CNA)

Cohesion Network Analysis (CNA; M. Dascalu, Trausan-Matu, McNamara, & Dessus, 2015) combines advanced NLP approaches with Social Network Analysis (SNA; Scott, 2017) to analyze and provide an in-depth view of discourse structure centered on text cohesion. SNA represents and examines social structures using graph theories; CNA is closely correlated to SNA as it provides equivalent indices to evaluate participation using network graphs that use estimates of discourse cohesion to simulate information exchanged between participants. CNA improves SNA because it considers semantic cohesion based on NLP when modeling students' interactions. As a consequence, CNA considers both students' interactions and discourse content to conduct an in-depth analysis of students' interactions.

Within the CNA, cohesion is computed using various similarity measures from different semantic models, namely: Latent Semantic Analysis (LSA; Landauer & Dumais, 1997), Latent Dirichlet Allocation (LDA; Blei, Ng, & Jordan, 2003) or word2vec (Mikolov, Chen, Corrado, & Dean, 2013). The cohesion graph represents a multi-layered structure, consisting of different nodes and the links between them, and it can be used as a proxy for the semantic content of the discourse (M. Dascalu, 2014). The cohesion graph consists of a central node, which represents the conversation's thread. Then, the central node is divided into contributions, which are further divided into sentences and words. Links are built to compute a cohesion score that denotes the relevance of a contribution in a conversation, or the impact of a word in a sentence or contribution. The graph also includes explicit links added by the participants in the conversations, such as "reply-to". Besides predicting collaboration (M. Dascalu, McNamara, Trausan-Matu, & Allen, 2018) and course grades (M. Dascalu, McNamara, Trausan-Matu, & Allen, 2018), CNA was also successfully employed to predict blogger community response to newcomer inquiries via automated dialog assessment (Nistor, Dascalu, Serafin, & Trausan-Matu, 2018; Nistor, Dascalu, Tarnai, & Trausan-Matu, 2020).

Comprehension is a difficult and challenging process, for which learners need to understand words and sentences, connect ideas, and link them to prior knowledge, while creating a coherent mental representation of the read text. One important factor in the comprehension process regards the cohesion of text (McNamara, 2004), which considers the degree to which there are semantic links between ideas within a text. Cohesion is higher when there are multiple ideas and words that overlap and when the connections between ideas are explicit. Low cohesion text is more challenging to understand, particularly for low knowledge and less skilled readers (O'Reilly & McNamara, 2007). The process of overcoming cohesion gaps is even more challenging when learners are faced with multiple documents that require establishing connections both within and between disparate text fragments. Making connections across multiple texts is considerably more difficult than doing so within a single text. Some text fragments may be semantically linked, while others may be isolated, distal, and thus more difficult to recognize or infer.

Part 2 – Empirical Studies

5 Modelling Writing Style through Textual Complexity Indices

This chapter focuses on two case studies centered on investigating the writing style in Romanian media discourses on the economic crisis and military discourse.

5.1 Case studies

The first case study (Terian, Cotet, Sirbu, Dascalu, & Trausan-Matu, 2019) analyzes discourses from Romanian media about the economic crisis spanning the years between 2008 and 2018. Not many people are specialists in economics, but everyone tends to have the impression of knowing something about economic crises—and, above all, is eager to talk about it. The economic crisis seems to have been the dominant topic of the last decade at least in Romania, generating a plurality of parallel or even contradictory discourses. Therefore, the usage of computerized tools for the analysis of such texts is welcome.

The second case study (Dragomir, Dascalu, Dascalu, Terian, & Trausan-Matu, 2020) explores military discourse from a diachronic perspective focused on an in-depth linguistic analysis of the differences between the writing styles of NATO official documents. The approach consists of a quantitative investigation of the evolution of this specific language throughout two different periods – the Cold War (1949 – 1990) and the post-Cold War period (1991 – 2018) –, with explicit emphasis on the examination of discourse corresponding to power dynamics. The main aim of the study is to describe and explain the manner in which two predominant power relations – integrative and adversarial – have been reified in the Alliance's discourse over a seventy year time span, tightly correlated with the historical and political context.

5.2 Research Question

The two case studies mentioned in section 5.1, address the following research question:

- **RQ1:** To what extent are textual complexity indices predictive of differences in writing style when performing multi-lingual analyses on different corpora and discourse types (i.e., NATO official documents and Romanian media)?

5.3 Method

Economic Crisis reflected in Romanian Media - Corpus

The corpus for this case study included 200 texts, grouped in 4 categories (specialized academic publications, specialized non–academic publications, non–specialized non–academic publications, informal debate), equally covering – in terms of amount (50), not necessarily length-wise. All selected texts were published between 1 January 2008 and 31 December 2018. We aimed to cover all the eleven calendar years included in this timeframe, but it was not our intention to obtain an equal distribution of the texts by year, as such a homogenization would be artificial. Instead, as far as the last two categories are concerned, we sought to include the sources from which we could single out texts that cover a wide range of ideological perspectives (at least from the viewpoint of the Left versus. Right opposition).

NATO Discourses - Corpus

The corpus for this case study was selected from NATO's archives, available online as public documents. The selection of the corpus was managed by taking into consideration a few important criteria: the type of power investigated (integrative versus adversarial), the time span (1949 – 2018), the relevance of pertinent texts carefully chosen from a comprehensive collection of documents (more than 1,000), the size and accessibility of the archives (some of which are outdated or contain missing links), and the topical variety of the texts (over 25 thematic categories and more than 150 topics). The empirical material used as basis for the analysis is primarily composed of NATO official documents resulted from 114 Ministerial Meetings (63 at the level of the Ministers of Defense and 51 at the level of Foreign Ministers) and 30 Summits, which occurred between 1949 and 2018.

Multi-lingual Textual Complexity Indices

The textual complexity indices for Romanian Language (Romania Media case study) and English Language (NATO case study) were generated using the *ReaderBench* framework (M. Dascalu, 2014; M. Dascalu, Dessus, Bianco, Trausan-Matu, & Nardy, 2014). Taking into account the dimension of our corpus for Romanian language which surpasses 800 million words extracted from online public sources, only the word2vec model was used in the current experiments, besides WordNet (Miller, 1995) semantic distances. Around 200 complexity indices tailored for Romanian language were generated using the *ReaderBench* framework and were used to exploring differences between four writing styles, all addressing the economic crisis. The textual complexity indices covering lexical, semantic, and cohesive features were used in statistical analyses to highlight differences in writing style between the selected documents. Regarding the NATO case study, more than 800 complexity indices, including all word-list indices, were computed using the *ReaderBench* framework. The underlying semantic models were trained using the Corpus of Contemporary American English (COCA) corpus (Davies, 2010).

Statistical Analyses

The *ReaderBench* framework facilitates a thorough statistical analysis of various properties including lexicon, syntax, semantics, text cohesion, and discourse structure. The indices generated by the *ReaderBench* framework (for both case studies) were further filtered on different criteria, as described below.

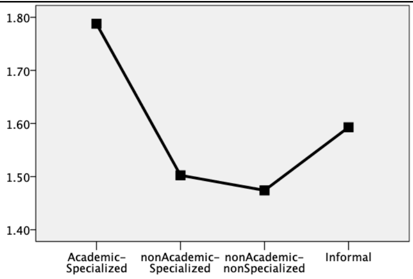
The first criterion was linguistic coverage, namely that an index is relevant and can be computed for at least 20% of the documents; indices with low linguistic coverage were disregarded. Most of those indices were related to the number of specific word lists. Second, the normality of the distribution of each index was checked; more precisely, indices with the absolute value of skewness or kurtosis greater than 2 were eliminated. Third, multicollinearity tests based on pair-wise comparisons were performed ($r > .90$) and only the most predictive indices were retained. The fourth criterion consisted of Levene's test of equality of error variances and disregarding indices whose resulting p-values are significant ($p < .05$).

For the Romanian Media case study, statistical analyses were conducted to examine differences in writing styles across the 4 document types (specialized academic, specialized non – academic, non – specialized non – academic, informal). For the NATO case study, statistical analyses were conducted to investigate differences in the writing styles of NATO discourses based on the type and time period (2 intervals between 1949 and 2018) in which they were produced. Our analysis focused on the lexical, semantic and cohesive properties of the analyzed documents. One-way ANOVAs for each textual complexity index were conducted to identify significant differences in writing styles. Afterwards, a stepwise Discriminant Function Analysis (DFA) (Klecka, 1980) was performed on both case studies.

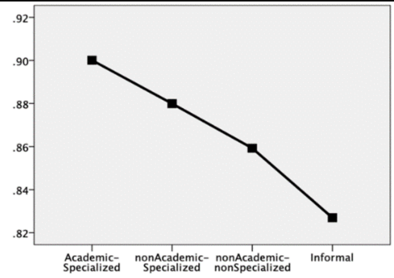
5.4 Results

Discourse Differences in Romanian Media

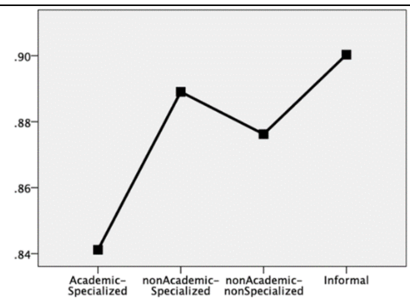
The analysis started from approximately 200 textual complexity indices for Romanian language which were further filtered on different criteria, as described in the Statistical Analyses section. After the four filters were sequentially applied, 17 indices passed, and were subject to one-way ANOVAs to identify significant differences in writing styles across the 4 document types. The most representative 8 indices in terms of fluctuations are displayed in Figure 2.



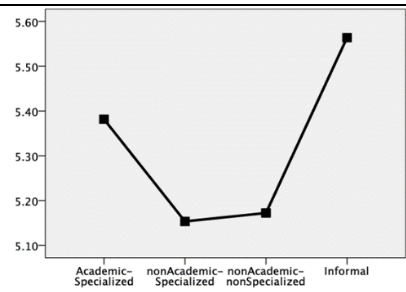
(a) Word length difference between stem and original form (M)



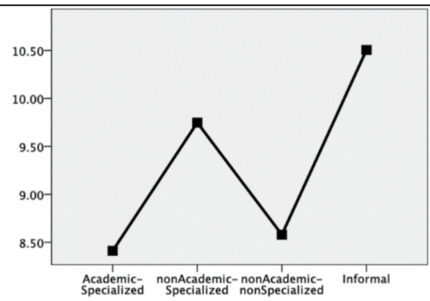
(b) *Local cohesion* – Wu-Palmer distance between sentences and paragraphs (M)



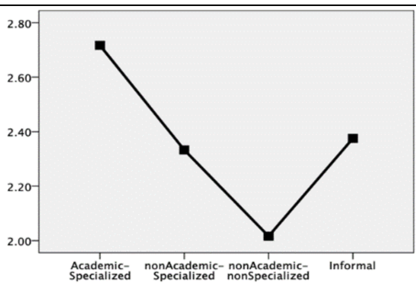
(c) *Global cohesion* – word2vec similarity between paragraphs and document (M)



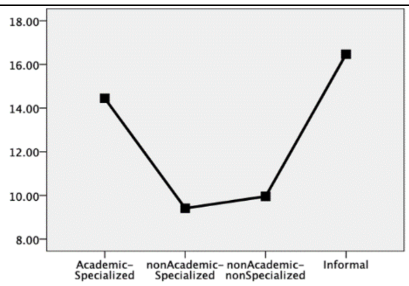
(d) Word entropy



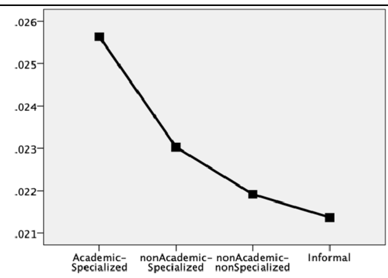
(e) Unique content words per sentence (M)



(f) Coordinating conjunctions per sentence (M)



(g) Paragraph CNA score (SD)



(h) Percentage of words included in lexical chains

Figure 2. a-h: Comparison between writing styles in terms of textual complexity indices.

A stepwise DFA was performed in order to predict the document type based on its underlying writing style properties. DFA reported an accuracy of 71.05% for correctly categorizing 143 documents (48+27+29+39) from the total of 200. Similar results were obtained for the LOOCV (Leave – One – Out Cross – Validation) which reported an accuracy of 64.50%. All F1-scores exceed the double of random baseline, however the differences in F1 scores for the 4 categories are considerable. A clear distinction can be noted between academic specialized documents and all other types (F1-score of 96%), followed by informal texts (F1 score of 67%), whereas the classifier has the biggest problems in identifying differences between specialized and non-specialized non-academic works, which are similar in writing style.

Adversarial versus Integrative NATO Discourses

First, 91 of the textual complexity indices computed by the *ReaderBench* framework were eliminated due to a low linguistic coverage, as the underlying text elements were not frequently encountered in NATO discourse (i.e., they were not present in at least 20% of all documents). Based on the specificities of the discourse, our corpus does not contain a significant number of occurrences of certain cue phrases or lexical dependencies.

Second, normality was checked in terms of Kurtosis and Skewness whose absolute values need to be below than or equal to 2; all variables exhibiting higher values were disregarded in follow-up analyses.

Third, all variables were tested using Levene's test of equality of error variances and those indices for which the resulting p-values are significant ($p < .05$) were disregarded as they exhibited a difference between the variances in the population. Thus, 341 indices were retained and were entered into two-way ANOVAs to examine whether the documents' properties differ across document's type (integrative and adversarial) and period (1949 – 1990 and 1991 – 2018). In order to better understand the differences, we present in Figure 3 the profile plots computed using the estimated marginal means for the most representative textual complexity indices. These values, corroborated with their visual representation from Figure 3, are the foundation for the detailed discussions from the following section. The most predictive indices explained considerably more variance in terms of period (partial $\eta^2 = .473, p < .001$) in contrast to type (partial $\eta^2 = .208, p < .001$), denoting that there were higher differences in time than between types.

Afterwards, a stepwise DFA was performed to predict the type and period of a given text based on the underlying writing style properties. All remaining variables were removed and considered non–significant predictors. The results prove that the DFA using these five indices significantly differentiated texts, Wilks' $\lambda = .849$, $\chi^2(\text{df} = 3) = 29.301, p < .001$. The DFA correctly allocated 104 (21+21+23+39) of the 184 documents from the total set, resulting in an accuracy of 56.50% (the chance level for this analysis is 25%). For the LOOCV, the discriminant analysis allocated 101 (20+20+23+38) of the 184 texts for an accuracy of 54.90%.

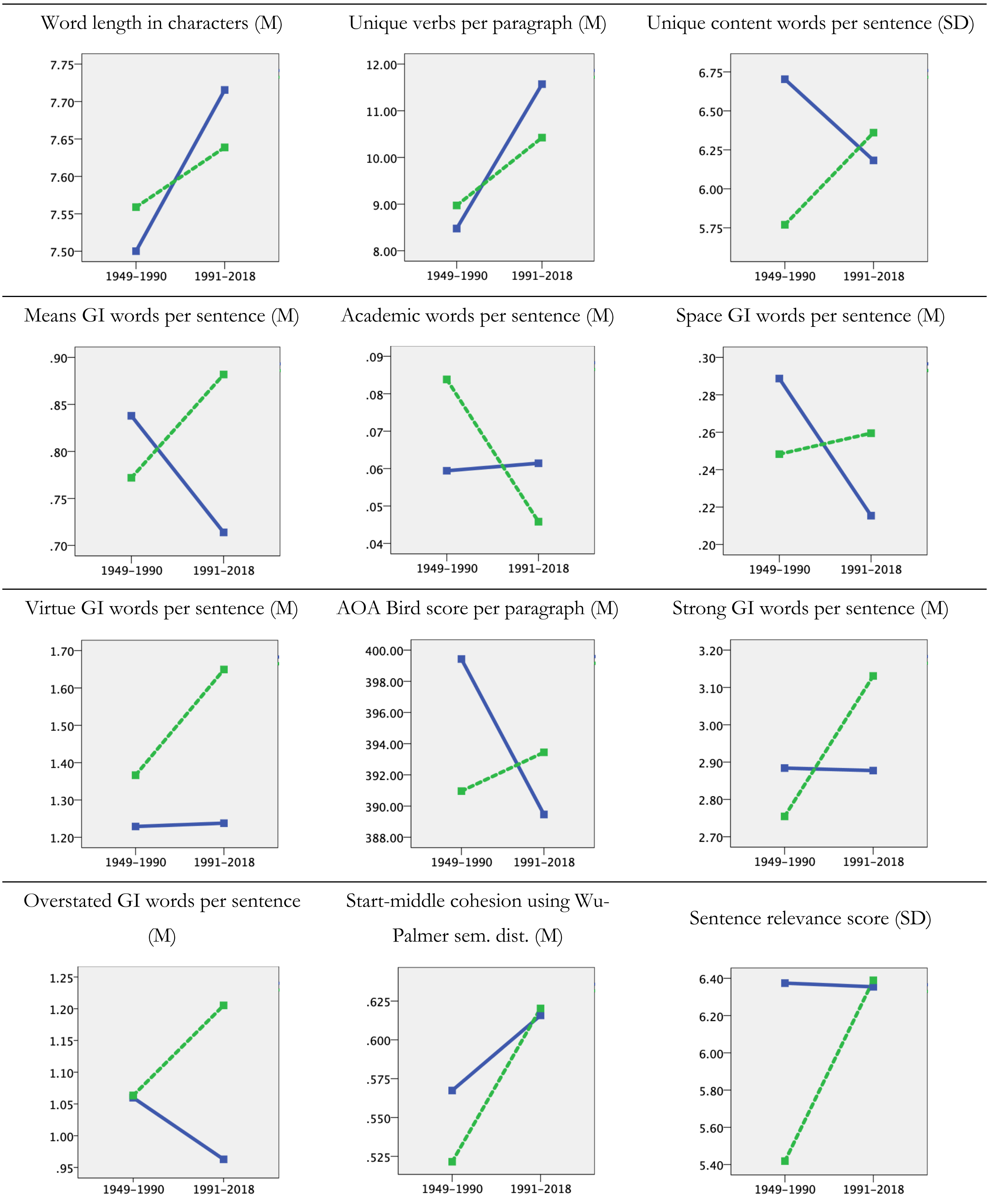


Figure 3. Profile plots based on the estimated marginal means of each textual complexity index.

(Adversarial type is blue, whereas Integrative is dotted green)

6 Assessing Online Student Performance through CNA

This chapter focuses on two case studies centered on assessing online student performance through CNA.

6.1 Case studies

The first study examines math success within a blended undergraduate course, which consisted of face-to-face lectures and support from online tools including a standard question-answer Piazza forum (Crossley et al., 2018). Our method is based on CNA variables and click-stream data to predict math success. Moreover, different sociograms (e.g.., interaction graphs between participants) were generated through CNA in order to analyze students' involvement and interactions, and to identify temporal trends between students (Sirbu et al., 2018a). In addition, a longitudinal analysis was performed (Sirbu, Dascalu, Crossley, McNamara, & Trausan-Matu, 2019), and different views were generated to model trends in student participation, as well as concept maps relying on the semantic relatedness between keywords found in the discussion forums.

The second study evaluates online activity of students in a Romanian Moodle course, by analyzing students' behaviors, modeling their interaction and predicting students' grades based on their online participation (M.-D. Dascalu et al., 2021). A Recurrent Neural Network with LSTM cells that combines global features, including participation and initiation indices, with a time series analysis on timeframes is used to predict student grades, while multiple sociograms are generated to observe interaction patterns. Students' behaviors and interactions are compared before and during COVID–19 using two consecutive yearly instances of an undergraduate course in Algorithm Design, conducted in Romanian using Moodle.

6.2 Research questions

The two case studies mentioned in section 6.1, addressed the following research questions:

- **RQ2:** To what extent do features generated by the *ReaderBench* framework, including CNA applied on forum discussions, textual complexity indices, longitudinal analysis, and features derived from click-stream data, predict student performance?
- **RQ3:** What are the differences in students' behaviors and interactions before and during the COVID-19 pandemic in two consecutive yearly installments of an undergraduate course?

6.3 Method

This chapter presents the overall method used for both studies: modeling success in a math course and student performance in Moodle courses.

Modeling Success in an Online Math Course - Corpus

For this case study, we used the data from a discrete math course for undergraduate students in a Computer Science Department (Crossley, Barnes, Lynch, & McNamara, 2017). The course was blended and included standard lectures (face-to-face), office hours, and support from online tools. The tools included a standard question-answering Piazza forum for students, teaching assistants, and instructors.

Student Performance in Moodle Course - Corpus

For this case study we collected data from two different Moodle instances from the 2018-2019 (*normal* conditions) and 2019-2020 academic years (*COVID-19* pandemic conditions) on a Moodle course from the second semester, held in Romanian and centered on Algorithm Design. The data included forum posts of students, lecturers and teaching assistants, and their online activities extracted from click-stream log data. The information collected from forum posts consisted of usernames, contributions' timestamps, reply-to links, and the actual texts from the posts.

Processing pipeline

Our approach is grounded in CNA combined with Machine Learning techniques, and it is used to evaluate and model students' participation and interactions. In contrast to the previous studies performed by M.-D. Dascalu et al. (2020), M. Dascalu, McNamara, Trausan-Matu, and Allen (2018) and Crossley, Paquette, Dascalu, McNamara, and Baker (2016), we introduce an integrated pipeline that accounts for all types of indices (CNA, time series and textual complexity), actions derived from clickstream logs, and a neural network for predicting course grades. We consider students' behaviors as revealed by actions within the OLE (e.g., viewing assignments, completing assignments, posting comments), as well as social interactions and the semantic content of their online contributions.

Figure 4 presents the automated processing pipeline which includes two important stages. The first stage consists of an ETL (Extract Transform Load) process that starts with the collection of click-stream data and forum discussions from the Moodle platform which were exported from the relational database (e.g., MariaDB). The second stage is centered on the automated processing pipeline from the *ReaderBench* framework, which has as input data the two datasets generated in the first stage.

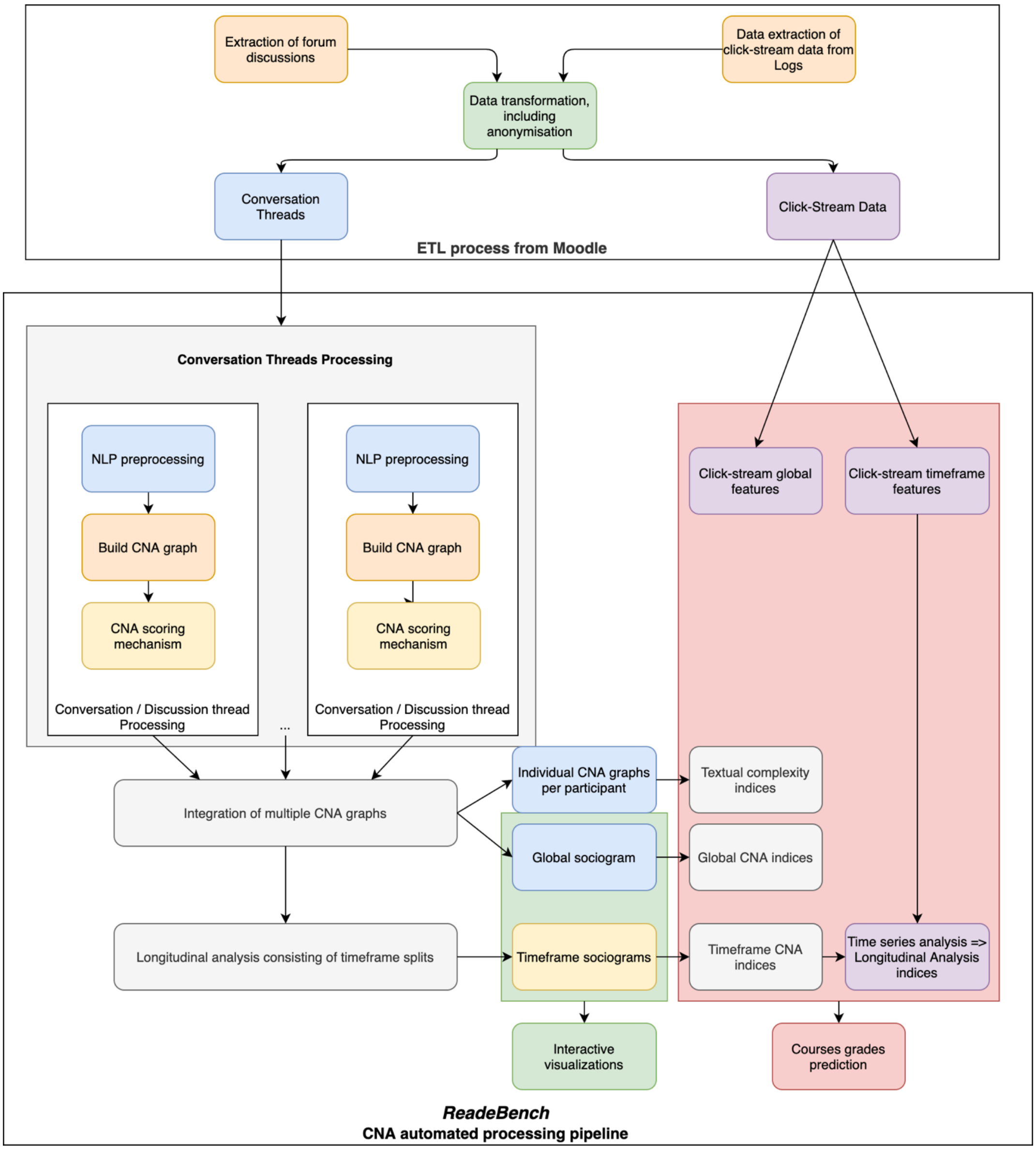


Figure 4. *ReaderBench* – Processing pipeline for predicting grades and generating interactive visualizations.

Predicting Course Grades

All indices including CNA, textual complexity, longitudinal analysis, as well as time series applied on timeframes and click-stream data are used to predict students' course grades using various machine learning algorithms. All previous indices provide valuable insights in terms of participation and collaboration (CNA indices), stylometry (textual complexity indices), online activity (click-stream data), as well as regularity and evolution in time (longitudinal analysis).

Our model consists of a Recurrent Neural Network (RNN) with LSTM cells (Hochreiter & Schmidhuber, 1997) that combines global features (CNA, textual complexity, and LA indices) with a time series analysis on timeframes in order to predict student grades. In our case, the inputs are weekly student activities derived both from CNA indicative of online participation and click-stream data reflecting overall interactions with Moodle.

Generating Interactive Visualizations

Interactive visualizations are generated using the global and timeframe sociograms, to highlight the interaction between participants and to depict the evolution of the community from one week to the next. Thus, multiple types of visualization are rendered to depict students' evolution, behaviors, and interaction patterns using the d3.js library (Bostock, 2021). The web application was built using Angular 6 framework, while various JavaScript libraries were integrated to create the interactive sociograms.

6.4 Results

Predicting Success in Online Math Course

The hierarchical edge bundling perspective in Figure 5 shows the interaction in a radial manner where dependencies are grouped into spline bundles and participants are grouped into their corresponding cluster/layer. Central participants are colored in blue, active members are displayed in green, and peripheral members in orange. On mouseover, the user can see the incoming and outcoming edges. Incoming edges and corresponding nodes (dependents) are displayed in dark blue, while outgoing links and outbound nodes (dependencies) are colored in red (see Figure 5 for participant ID 303190984).

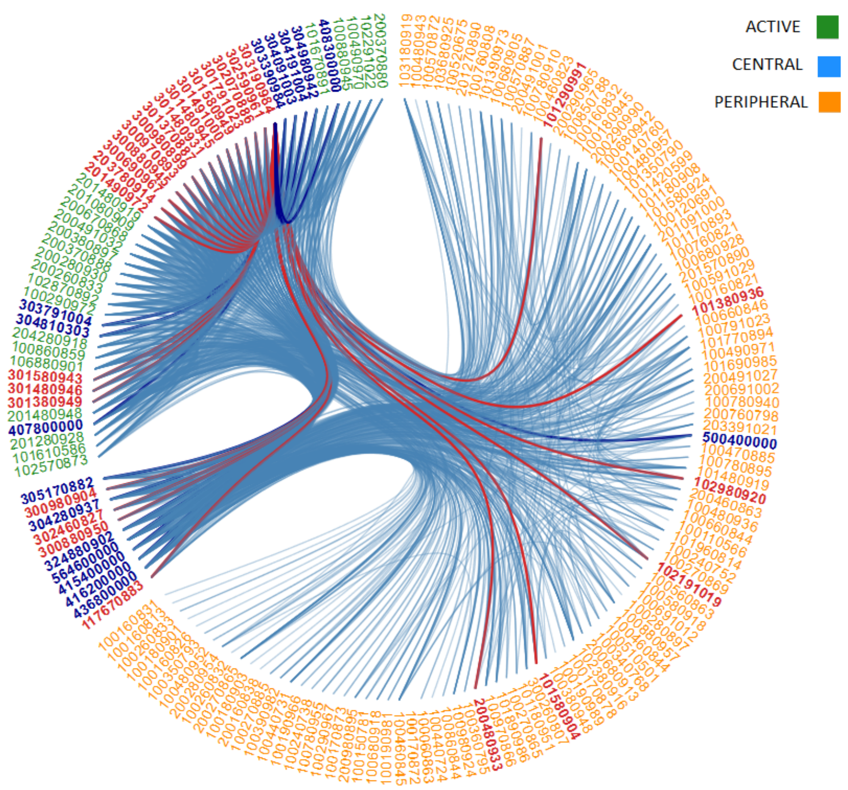


Figure 5. Global sociogram for the entire course period (Aug 23rd – Dec 24th, 2013).

Figure 6 shows a *force-directed graph* that considers the strength of the communication between nodes (i.e., students). The width of the edges is proportional to the text quality from CNA (i.e., cumulative contribution scores of exchange messages), whereas the length of each edge is automatically rendered by the visualization library. In addition, each node's size is proportional to the average indegree and outdegree scores of each participant. Weekly snapshots are generated to examine how participation in the community evolves from one week to the next.

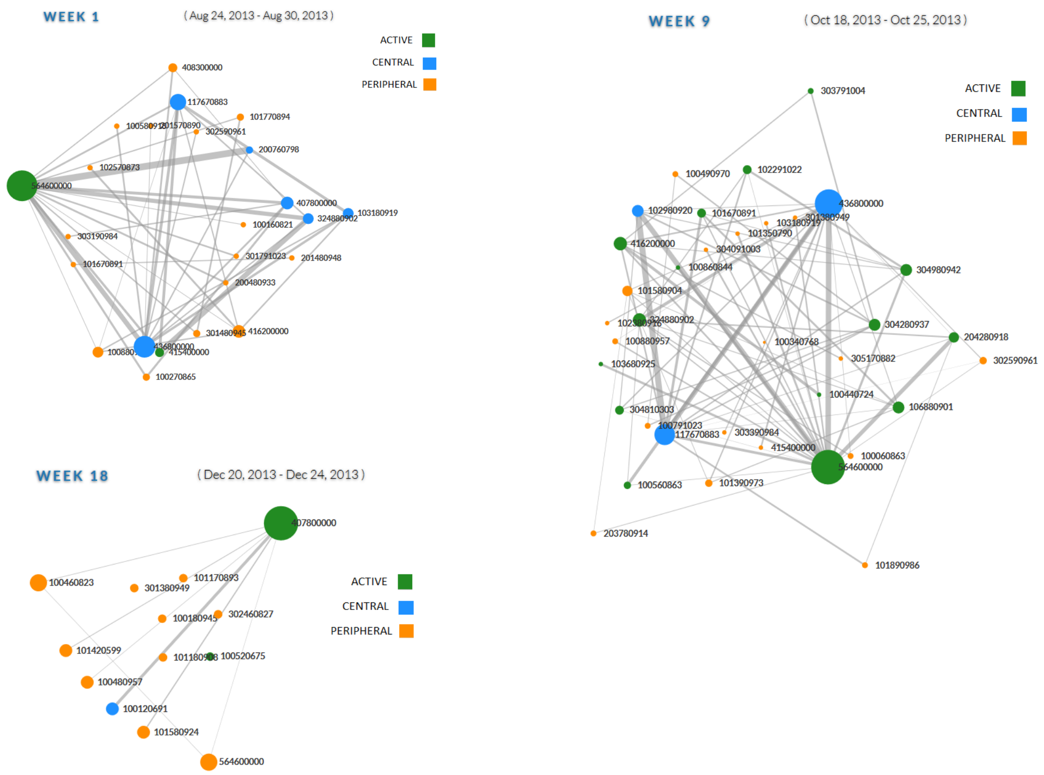


Figure 6. Weekly snapshots of course sociograms.

Specific traits of the behavioral interaction patterns can be observed from the two types of visualizations, namely: a) a dominance of peripheral members having the lowest number of interactions; b) a growing degree of collaboration from the peripheral layer to the active members, and more importantly, to the central participants (including the course instructors and teaching assistants); c) a rather slow start of the community (week 1), followed by an increase in participation (week 9) and a drastic decrease in the last week; and d) although directed towards more central participants, free discussions can be observed and the course is not dominated by a single individual, a situation common in many MOOCs and online communities.

The click-stream and the CNA variables explained around 20% of the variance and indicated that higher math scores were best predicted by students who had lower recurrence (i.e., distance expressed as number of weeks between two consecutive timeframes with non-zero values) of postings that generated collaborative effects (both Social KB and CNA indegree) with other participants. If participants regularly contribute to each consecutive timeframes from the longitudinal analysis, their corresponding recurrence score is zero; in contrast, if additional successive weeks are skipped, their recurrence scores increase. In contrast to CNA scores that denote active involvement, recurrence quantifies unbalance and inconsistent participation over time.

Predicting Course Grades in Moodle

There were many students who were isolated on the forum in the first academic year, without interacting with peers or responding to other student inquiries. Also, only two teaching assistants were more active in all conversations, whereas lecturers were less engaged. In contrast, students were more engaged in the discussions from the 2019-2020 academic year, collaborated more, and there were fewer students who solely introduced isolated posts as new conversation threads.

In addition to analyzing students' behavior and interactions, we examined which keywords were most frequently used in forum discussions. Some concepts' usage depended on the week (with or without homework deadlines), whereas others were frequently used throughout the entire academic year. In the weeks with homework deadlines (i.e., weeks 8 and 13) , keywords like "lucru" (eng. "work"), "problemă" (eng. "problem"), "merge" (eng. "run") were also encountered (see Figure 7). Compared with the previous academic year, keywords like "problem" and "work" were intensively used during the entire course, not only in the weeks with homework deadlines.

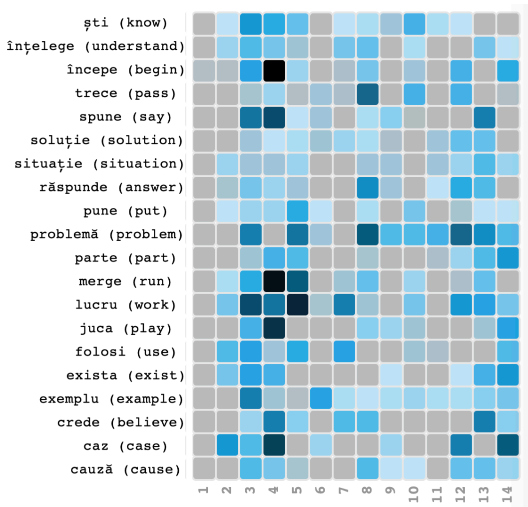


Figure 7. Concept Map 2020 – The most discussed topics.

We explored differences in students' behaviors and interactions before and during COVID–19 using the CNA, textual complexity, and longitudinal analysis indices that entered the analysis after normality and multi-collinearity checks. As expected, a significant increase in online activity was observed (e.g., higher CNA contribution scores) during the 2019-2020 academic year impacted by

COVID-19. In addition, considerably more threads were initiated, and the overall network is more connected (e.g., higher CNA eigenvector values).

Textual complexity indices denoted a more elaborated and sophisticated discourse in the second academic year. Longitudinal analysis indices also revealed interesting results. An increased activity towards the end of the semester was observed in the 2018-19 academic year (e.g., the overall slope is positive). In contrast, the COVID–19 outbreak generated an off-balance, a drastic increase in participation, that afterwards decreased towards the end of the semester (e.g., the slope was close to zero or negative, with higher standard deviations).

Models were trained individually for each academic year. The models for both academic years were trained for a maximum of 2000 epochs with early stopping based on RMSE on the training loss, with a patience of 200 epochs. A dropout layer of 0.2 was introduced before the last hidden layer to reduce overfitting and improve generalization. These parameters were chosen based on the performance obtained on cross-validation.

For the 2018-2019 academic year, the model that includes all components except for the CNA weekly features and using an LSTM cell size of 24 and a hidden layer size of 16, explained the largest percent of variance (R^2 = .27). For the subsequent year, during COVID-19, the best performance was obtained with the same features, but with a hidden layer size of 24 instead of 16. This model obtained a higher R^2 of .34. Global CNA indices can be perceived as sums of weekly CNA indices from the longitudinal analysis; thus, it is not really surprising that they are not always useful in the RNN architecture. A higher explained variance in the second academic year is justifiable, given the larger amount of collected data (e.g., more posts and a denser network) that helps create a more predictive model. In both academic years, the removal of textual complexity indices leads to the largest decrease in explained variance.

7 Exploring Dialogism using Language Models

This chapter focuses on two case studies centered on exploring dialogism using Language Models.

7.1 Case studies

The first case study introduces a method grounded in dialogism for evaluating students' involvement in chat conversations based on semantic chains computed using language models. These semantic chains reflect emergent voices from dialogism that span and interact throughout the conversation.

The second case study applies our model on an automated essay scoring task. Our goal was to evaluate the prediction accuracy of the features derived from the semantic chains on the task of automated text scoring.

7.2 Research question

The two case studies mentioned in section 7.1, addressed the following research question:

- **RQ4:** To what extend are semantic chains operationalized using language models predictive in both collaborative (i.e., evaluating collaboration between chat participants) and individual settings (i.e., automated essay scoring)?

7.3 Method

Measuring Involvement in CSCL Conversations - Corpus

Our analysis is performed on the same chat conversations processed in detail by Dascalu et al. (M. Dascalu, McNamara, Trausan-Matu, & Allen, 2018; M. Dascalu, Trausan-Matu, McNamara, & Dessus, 2015). This corpus consists of 10 chats selected from a corpus of more than 100 conversations which were scored by 4 raters. The conversations took place between four to five undergraduate students studying Computer Human Interaction who debated on the advantages and disadvantages of specific CSCL technologies (Trausan-Matu, Dascalu, Rebedea, & Gartner, 2010).

Automated Essay Scoring - Corpus

We used the Automated Student Assessment Prize (ASAP) dataset (Kaggle Inc, 2021) to evaluate the semantic chains. The essay corpus consists of 8 different sets of essays, each with its own scoring scale, this all grades were normalized to 0-1 depending on their type. Each essay from ASAP dataset has a length of about 150-550 words.

Corpus for building Semantic Chains

In order to find the heads capable of detecting semantic links between words that belong to the same chain, a specific dataset containing examples of links needed to be compiled. A set of simple heuristics were used to extract links from sample texts, for all pairs of words tagged as noun, verb, or pronoun that check one of the following conditions: repetitions of words having the same lemma; synonyms, hypernyms, or siblings in the WordNet taxonomy; coreferences identified using spaCy. The TASA corpus (Zeno, Ivens, Millard, & Duvvuri, 1995) was selected as reference and the previous pairs of words were marked throughout the entire dataset.

Predictions of Semantically Associated Words

No single attention head is accurate enough to predict these kinds of semantic relationships between words. Therefore, a prediction model that learns to combine the attention values from all the attention heads between two words was trained on the dataset constructed based on TASA (Zeno, Ivens, Millard, & Duvvuri, 1995). By considering both directions of the attention heads, 288 scores were used in total, similar to the approach used by Clark, Khandelwal, Levy, and Manning (2019). Different architectures were trained and evaluated: a linear model that only computes one weight for each attention head, and multi-layer perceptron (MLP) with one or two hidden layers. All models return one number passed through a sigmoid activation.

The best performing models on this dataset are not necessarily the best ones to be used in follow-up analyses. Our goal was to detect pairs of words that belong to the same semantic chain. The target used to build this dataset is a very simple one and we do not want the attention model to perfectly predict it, but rather to be capable to generalize to other similar types of links. Besides this, a linear model is easier to interpret, which might provide an additional benefit. Therefore, this initial validation was primarily used to find the best parameters for different types of models, which are subsequently evaluated using a different task.

The prediction model previously described can be used to score all pairs of words that are within a given distance in the text. The next step consists of grouping these pairs of words into sets of semantically related words, i.e., semantic chains. In order to filter the links based on the predicted weight, a fixed threshold was used. The semantic chains are selected in the form of connected components from the resulting graph.

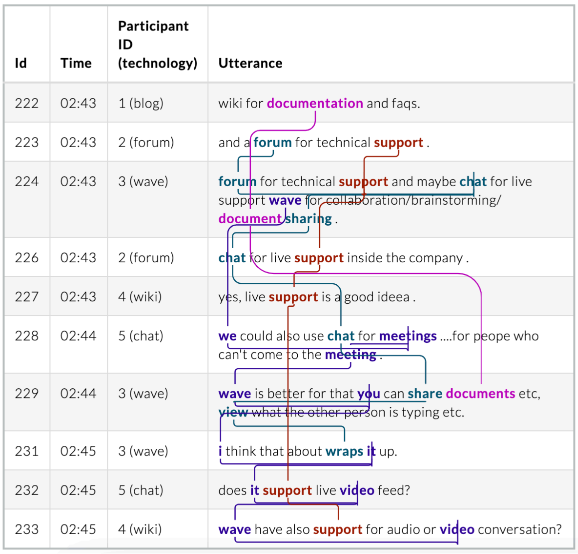
Building Semantic Chains

Our method uses contextual information captured by BERT (Devlin, Chang, Lee, & Toutanova, 2018) to identify the links in a semantic chain. The trained model was then used to identify all potential semantic links in a conversation, while accounting for all pairs of words. Starting from these links, semantic chains are generated as connected components in the graph obtained by interconnecting all links exceeding an imposed similarity threshold.

7.4 Results

Visualizing semantic chains

Interactive visualizations were introduced to highlight both a longitudinal propagation of voices (see Figure 8) and transversal overlap of semantic chains between participant (see Figure 9). The views were developed using Angular 6 (Google Inc, 2010-2021), while the links between words were drawn using SVG. The same chat excerpt from Figure 8 was analyzed by M. Dascalu, Trausan-Matu, McNamara, and Dessus (2015). Words are colored according to the semantic chain two which they belong as well as the corresponding links. Each row represents an utterance, enriched with following details: identifier, timestamp, and participant identifier, which was incrementally generated for anonymization and followed by the supported technology by each participant.



Id	Time	Participant ID (technology)	Utterance
222	02:43	1 (blog)	wiki for **documentation** and faqs.
223	02:43	2 (forum)	and a **forum** for technical **support** .
224	02:43	3 (wave)	**forum** for technical **support** and maybe **chat** for live support **wave** for collaboration/brainstorming/ **document** **sharing** .
226	02:43	2 (forum)	**chat** for live **support** inside the company .
227	02:43	4 (wiki)	yes, live **support** is a good ideea .
228	02:44	5 (chat)	**we** could also use **chat** for **meetings** ....for peope who can't come to the **meeting** .
229	02:44	3 (wave)	**wave** is better for that **you** can **share** **documents** etc, **view** what the other person is typing etc.
231	02:45	3 (wave)	**i** think that about **wraps** **it** up.
232	02:45	5 (chat)	does **it** **support** live **video** feed?
233	02:45	4 (wiki)	**wave** have also **support** for audio or **video** conversation?

Figure 8. Longitudinal view of semantic chains within a conversation.

In contrast with the initial findings of M. Dascalu, Trausan-Matu, McNamara, and Dessus (2015), our method identified more semantic chains with more related words – for example, new semantic chains – concepts related to documentation and actions; more related words – "meetings" and "video" are related to "wave"; chats and forum are now aggregated together as language models grasps their similarity. Figure 9 presents a transversal view of the occurrences of semantic links whose underlying concepts are uttered by different participants. The words and the links are colored according to their corresponding semantic chain; colors differ between the visualizations because they are randomly selected.

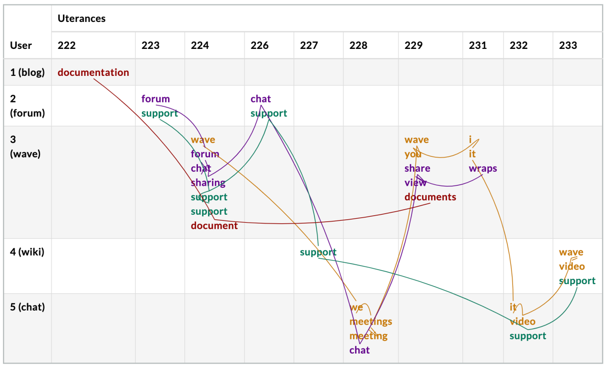
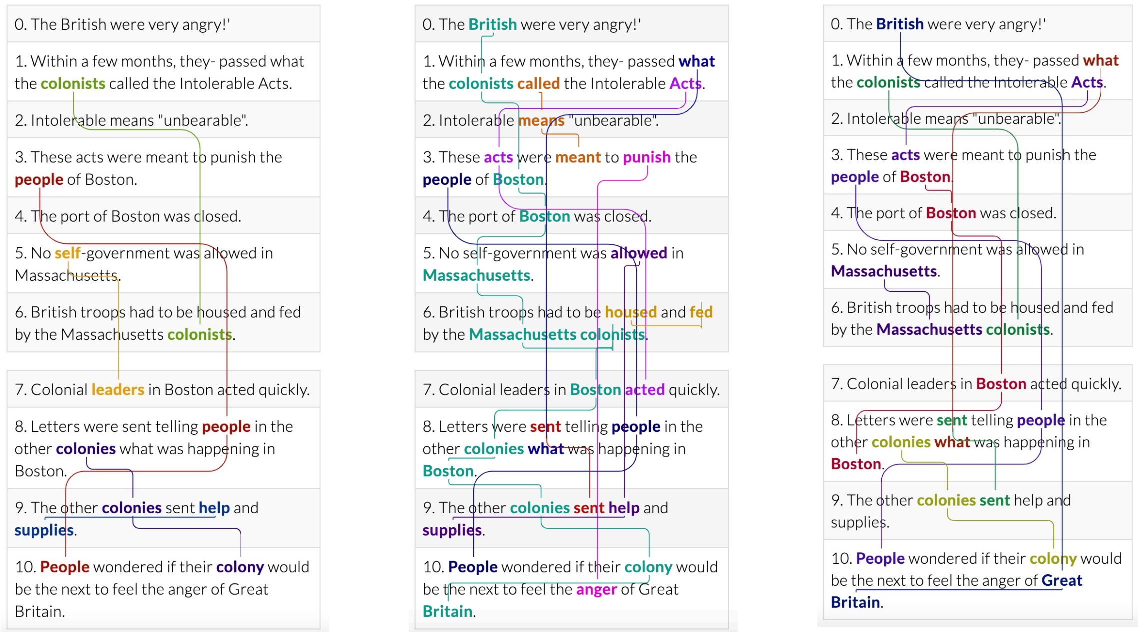


Figure 9. Transversal view of semantic chains within a conversation.

Moreover, an interactive view was introduced to visualize the semantic chains within a document that is divided in paragraphs, and subsequently in sentences (see Figure 10). Each sentence represents a row, while rows are grouped in their corresponding paragraph. Words that belong to different semantic chains are colored accordingly to their chain, while words from a single semantic chain are linked using a path having the same color as the chain.

First, we can observe the high density of the chains extracted with our method when compare to the classical lexical chains. In addition, surprising relations that were not present in the constructed dataset can be seen in the generated chains. The linear model found connections between ”colonists” and ”Boston”, or between ”help” and ”supplies”, while the MLP model identified connections between ”British” and ”Great Britain” as a compound word.



(a) lexical chains

(b) semantic chains using the linear model

(c) semantic chains using the MLP model

Figure 10. Visualizations of lexical and semantic chains.

Involvement Prediction in CSCL

Several machine learning algorithms were tested using the previously introduced features to evaluate the performance of semantic chains in assessing involvement of students within the conversations. Since the dataset contained 10 distinct conversations, a 10-fold cross-validation was performed, leaving one chat out for testing in each fold. Two different grades, namely participation (e.g., reflective of active involvement) and collaboration (e.g., interactions with peers) manually scored between 1 and 10 were predicted with two separate models.

The human performance obtained on this dataset shows a MAE lower than 1 out of 10, denoting that raters were close one to another, but did not perfectly agree among themselves (inherently, part were more relaxed while the other were more fastidious); as such, 4 ratings were gathered for each participant and the system predicts the raters' average rating for both participation and collaboration. Random forest is the most predictive model reaching a MAE of 0.55; lowering the number of generated trees has a beneficial impact on performance, given the limited number of features and examples. Overall, the machine learning models seem to better capture the average scores of the raters, having the advantage of generalizing across all participants.

A subsequent analysis was performed to understand the importance of each feature for the two different tasks.

Predicting Essay Scores

Semantic chains cannot be directly evaluated in the absence of an annotated dataset, so we decided to indirectly evaluate them on another task. Our assumption was that semantic chains can better reflect text cohesion and the quality of written essays. Therefore, we used the Automated Student Assessment Prize (ASAP) dataset (Kaggle Inc, 2021) to evaluate the semantic chains. Given that the validation and test partitions of the dataset are no longer available, the models were evaluated with 5-fold cross-validations on the training partition. To reduce the variance in the results and the influence of initial weights, the scores were averaged over 10 different runs.

Several features were defined based on the extracted semantic chains, some of them inspired from the LEX-1 set (Somasundaran, Burstein, & Chodorow, 2014). The goal was to evaluate the prediction accuracy of the features derived from the semantic chains and descriptive of text cohesion on the task of automated text scoring. Our features are descriptive of both local and global text cohesion as semantic chains span across the text, but also of semantic flow (O'Rourke & Calvo, 2009) while considering the recurrence and overlap of voices.

Different link prediction models and weight thresholds were compared with a baseline model relying on the Galley and McKeown (Galley & McKeown, 2003) lexical chains construction algorithm. The same features and classification algorithm were used for all chain discovery methods. Therefore, the differences in the performance of the models should only depend on the quality of the extracted semantic chains.

A simple neural network with a single hidden layer of size 16 was trained using the features previously described to predict the grade given to essays. Several threshold values for filtering the word pairs were visually inspected. Using small thresholds leads to large, connected components and do not separate well the voices in text, so only the results obtained with large thresholds are presented. The best mean average error (MAE) was obtained with the MLP model, while the linear model obtained the best R2 score; both models used a 0.9 threshold. However, the models account just for roughly 20% of variance; this is understandable, since flow is only part of what is in an essay.

8 Discussion

8.1 Identify Advantages of our Approach

The case studies presented in this thesis and incorporating the three objectives presented in Chapter 1.2, are based on advanced NLP techniques, and focus on the following tasks: identify and model writing style in texts, evaluate students' performance on online learning environments, identify interaction patterns between students, predict student's grades based on their online activity, identify semantic chains within texts, evaluate collaboration based on semantic chains in chats conversations.

To emphasize the advantages of our approaches, we present for each research question from Chapter 1.2, the proposed methods and the results obtained.

- **RQ1:** To what extent are textual complexity indices predictive of differences in writing style when performing multi-lingual analyses on different corpora and discourse types (i.e., NATO official documents and Romanian media)?

Textual complexity generated by the *ReaderBench* framework provide both insights into numerical properties of various textual features, as well as text component characteristics such as readability, local or global cohesion, or word complexity. Chapter 5 presents two case studies which investigate the writing style in Romanian media discourses on the economic crisis and NATO official documents, based on textual complexity indices generated by the *ReaderBench* framework. The textual complexity indices generated for each case study were used to exploring differences between four writing styles. Moreover, the textual complexity indices covering lexical, semantic, and cohesive features were used in statistical analyses to highlight differences in writing style. Besides WordNet semantic distances we use word2vec semantic model.

Analyses of Variance were conducted to identify the textual complexity indices that exhibited the highest significant differences in writing styles. Afterwards, a stepwise Discriminant Function Analysis was performed on both case studies.

For Romanian media, statistical analyses were conducted to examine differences in writing styles across four document types (specialized academic, specialized non-academic, non-specialized non-academic, informal). For NATO documents, statistical analyses were conducted to investigate differences in the writing styles of NATO discourses based on the type and time period in which they were produced (integrative and adversarial). The results reveal statistically significant and interesting differences with regards to the degree of word elaboration (length and polysemy count), the number of unique verbs per paragraph, the Age of Acquisition (AoA) score per paragraph, as well as the number of words within multiple lists.

- **RQ2:** To what extent do features generated by the *ReaderBench* framework, including CNA applied on forum discussions, textual complexity indices, longitudinal analysis, and features derived from click-stream data, predict student performance?

Chapter 6 presents two case studies which focuses on predicting math success and students' grades based on their online participation in two online environments: MOOC and Moodle course. Our processing pipeline integrates Cohesion Network Analysis indices, longitudinal analysis indices, textual complexity indices and click-stream data to predict students' grades using various machine learning algorithms. All previous indices provide valuable insights in terms of participation and collaboration (CNA indices), stylometry (textual complexity indices), online activity (click-stream data), as well as regularity and evolution in time (longitudinal analysis).

Linear models indicated that math success was related to days spent on the forum and by students who more regularly posted in the online class forum.

- **RQ3:** What are the differences in students' behaviors and interactions before and during the COVID-19 pandemic in two consecutive yearly installments of an undergraduate course?

Using Cohesion Network Analysis indices, various interactive visualizations were generated to highlight the interaction between participants and to depict the evolution of the community from one week to the next. Multiple types of visualization (e.g., force-directed graph, hierarchical edge bundling, radar chart, parallel coordinates chart) are rendered to depict students' evolution, behaviors, and interaction patterns using the d3.js library. Moreover, we examined which keywords are most frequent in student's discussions, and we represent the concepts using a concept heatmap. All visualizations are presented in Chapter 6.

Due to the COVID-19 pandemic, the transition from physical to fully online was drastic, adaptation was required in a short amount of time, and smart learning environments eased this transition, while supporting both students and teachers. Since everything moved online, face-to-face class discussions transitioned to forums, chats, and video meetings. Chapter 6 presents an analysis of students' online activity in a Moodle course for two years, before and during the COVID-19 pandemic. Students' behaviors and interactions are compared before and during COVID-19 using two consecutive yearly instances of an undergraduate course in Algorithm Design, conducted in Romanian using Moodle. Students' behaviors and interactions are compared in terms of differences derived from our wide range of indices, while also considering comparative interactive views illustrating their interactions, and providing a qualitative analysis from the tutor's perspective. The results showed that the COVID-19 outbreak generated an off-balance, a drastic increase in participation, followed by a decrease towards the end of the semester, compared to the academic year 2018-2019 when lower fluctuations in participation were observed.

- **RQ4:** To what extent are semantic chains operationalized using language models predictive in both collaborative (i.e., evaluating collaboration between chat participants) and individual settings (i.e., automated essay scoring)?

Transformer-based models (e.g., BERT) provide insights regarding the importance and relations between words, by looking at the attention values. Thus, we introduced and evaluate a method which use the contextual information captured by BERT to find the semantic links, while only considering the attention weights from different heads (Chapter 7). The resulting model generalizes to multiple relations including repetitions, semantically related concepts from WordNet, as well as pronominal resolutions. In contrast to the previous results, our new method identifies more semantic chains with more related words and includes coreference resolution (e.g., pronoun resolution).

Using the semantic chains identified with contextual information captured by BERT, we tested several machine learning algorithms to evaluate the extent to which semantic chains are predictive of student involvement in chat conversations. The ratio of chains, denoting the coverage of semantic chains used by a participant in relation to the entire conversation, is by far the most predictive feature. Continuations, reflective of links between two different participants, coupled with counts and ratios of chains covered by the participant are the best predictors for collaboration. Moreover, our model was successfully employed for automated essay scoring. Our aim was to evaluate the prediction accuracy of the features derived from the semantic chains and descriptive of text cohesion on the task of automated text scoring.

8.2 Faced Problems and Provided Solutions

One potential limitation of the study regarding students' performance in Moodle course (Chapter 6.1.2) is the generalizability of the model. One factor to consider regards students who exhibited lurking behaviors and did not have any active posts. These students could not be included within the analyses as there were no textual traces to analyze. This factor may influence the accuracy of model predictions when considering the entire population of students. Thus, additional mechanisms centered on log analytics (e.g., access to specific posts, homework completion rates) need to be taken into account to build baseline models capable of generating predictions for this category of students. Moreover, the current experiments need to be applied on a larger timeframe and on additional courses in order to create generalizable models across different course topics and situations. Longer timeframes (i.e., more installments) and additional courses will be considered to create predictive models with a higher degree of generalizability, which is hard, if not impossible, to obtain with the changes induced by the pandemic. Nevertheless, the specificities of each course, with underlying topics, activities, and interaction patterns, need to be considered while building prediction models.

One limitation of our approach regarding exploring dialogism using Language Models (Chapter 7) is the small ratio of coreference links in the generated dataset, which is also visible in the constructed semantic chains. One way to improve this in the future will also balance the rules used to generate the dataset and include more.

8.3 Educational Implications

The generated sociograms presented in Chapter 6 can be used by teachers to follow the evolution of students in term of interactions, interactivity, and online participation, thus enabling them to intervene when they notice a decrease in participation or inactivity. Students' chances of passing or obtaining a better grade can be increased if teachers encourage them to be more engaged throughout the course. As many learning institutions moved to online and face-to-face interactions and activities were drastically reduced, a mechanism that keeps track of students' activities and their likelihood of obtaining good scores would be beneficial for both students and teachers.

Moreover, the visualizations can be used by instructors and researchers to better understand participation and collaboration within large-scale learning environments. Weekly snapshots can be used to better understand students' evolution throughout the course, while correlating their activities with specific course events (e.g., homework deadlines, tests, holidays, exam, etc.).

9 Conclusions

9.1 Personal Contributions

A Modeling Writing Style through Textual Complexity Indices

The following contributions on modeling writing style through textual complexity indices were made:

- Performed detailed analyses of writing style in Romanian media discourses on the economic crisis highlight significant differences in terms of linguistic features;
- Identified and compared linguistic features specific to Romanian media discourses on the economic crisis spanning the years between 2008 and 2018;
- Performed detailed analyses of military discourse from a diachronic perspective focused on an in-depth linguistic analysis of the differences between the writing styles of NATO official documents;
- Conducted a quantitative investigation of the evolution of NATO official documents language throughout two different periods – the Cold War (1949 – 1990) and the post-Cold War period (1991 – 2018) –, with emphasis on the examination of discourse corresponding to power dynamics;
- Described and explained the manner in which two predominant power relations – integrative and adversarial – have been reified in the Alliance's discourse over a seventy-year time span, tightly correlated with the historical and political context.

B Assessing Online Student Performance through CNA

The following contributions on assessing online student performance through CNA were made:

- Examined math success within a blended undergraduate course using CNA, while controlling for individual differences and click-stream variables;
- Evaluated students' behaviors and interaction patterns in an undergraduate course on Algorithm Design offered at University Politehnica of Bucharest, for both 2018 – 2019 and 2019 – 2020 academic years, each reflecting different external conditions and demands, before versus during the COVID–19 pandemic;

- Predicted student performance using a recurrent neural network model that considers time series analyses on timeframes, in addition to the CNA, textual complexity, and longitudinal analysis indices;
- Generated various sociograms to show the interaction between participants, based on CNA indices;
- Generated weekly sociograms to examine how the interactions between peers and with tutors evolve from one week to the next. These visualizations provide insights into students' behaviors in association with course events, such as deadlines, assignments, tests, and exams;
- Examined the linguistic properties of the online discussion forums using CNA;
- Identified temporal trends among students;
- Introduced a longitudinal analysis and detailed visualizations of participation useful for gaining a better understanding of how students interact during an online math course.

C Exploring Dialogism using Language Models

The following contributions regarding exploring dialogism using language models were made:

- Introduced and evaluated a novel method of identifying semantic chains using state-of-the-art language models from NLP, namely BERT;
- Generalized to multiple relations including repetitions, semantically related concepts from WordNet (e.g., synonyms, hypernyms, hyponyms, and siblings), as well as pronominal resolutions;
- Introduced a method grounded in dialogism for evaluating students' involvement in chat conversations based on semantic chains computed using language models;
- Evaluated the prediction accuracy of the features derived from the semantic chains on the task of automated text scoring;
- Introduced interactive visualizations to highlight both a longitudinal propagation of voices and transversal overlap of semantic chains between participants in chats;
- Introduced interactive visualization to see the semantic chains within a document that is divided in paragraphs, and subsequently in sentence.

D Additional contributions

In addition to the contributions mentioned above regarding empirical studies, numerous contributions have been made in various studies:

- Created a novel smart learning environment for generating personalized feedback for student writings for both English (Botarleanu, Dascalu, Sirbu, Crossley, & Trausan-Matu, 2018; Sirbu, Botarleanu, Dascalu, Crossley, & Trausan-Matu, 2018) and Romanian (Sirbu et al., 2018b) languages;
- Analyzed participants' interaction on two types of communities: OKBCs and gaming communities to observe important traits and specific interaction patterns (Sirbu et al., 2017);
- Predicted course grades based on students' participation and evaluate the interactions between participants and the impact of homework deadlines and tests in their online contributions through interactive sociograms in a Moodle course centered on the Usage of Operating Systems for freshman students (M. Dascalu et al., 2018; M.-D. Dascalu et al., 2020; Gutu-Robu et al., 2018);
- Performed a comprehensive analysis of a debate community from Reddit, which is focused on discussing politic ideas and beliefs related to socialism and communism, in order to model the interaction between participants, analyze user behaviors based on regularity measures, extract the most discussed topics, and predict users' ranks based on their participation (Fetoiu et al., 2020);
- Introduced text processing and information retrieval techniques for implementing an intelligent platform for the advanced search of Romanian drugs and interactions between them (M.-D. Dascalu et al., 2019; Ene, Sirbu, Dascalu, Trausan-Matu, & Nuta, 2019; Paraschiv et al., 2019);
- Built a Romanian knowledge base for drug administration (Nicula, Dascalu, Sirbu, Trausan-Matu, & Nuta, 2019);
- Expanded the structure of the CNA graph (M.-D. Dascalu, Ruseti, Dascalu, McNamara, & Trausan-Matu, 2020) with lexical overlap links of multiple types, together with coreference links to highlight dependencies between text fragments of different granularities. Introduce two visualizations of the CNA graph that support the visual exploration of intra-textual and inter-textual links.

9.2 Directions for Future Research

In terms of future work regarding assessing online student performance, our aim is to introduce in-class evaluations in which we assess the impact of using the automated tools during the academic year that enable timely reactions from the tutors, in contrast to a posteriori assessment. In addition, we want to introduce custom configurations that consider specific semantic models tailored for the topic of each course. This will afford applications beyond the course targeted in this study. Another overarching objective is to provide automated feedback to both students and teachers regarding student involvement in the course. To this end, the CNA plugin will be expanded such that Tutors will be able to periodically evaluate the evolution of their students, identify at-risk students, and take timely actions. Further modifications are envisioned wherein students can be incentivized to more actively engage and collaborate more with their peers. Ultimately, the objective is to maximize students' chances of success in the course.

In addition, we envision applying the dialogical model on other datasets and performing related analyses. For example, dialogism can be used to monitor student engagement in online courses and the introduced features can be employed to predict dropout or course grades. Even more, by evaluating the introduced semantic chains, discussions and corresponding contributions can be catalogued as being course-specific, administrative, or off-topic, an automated guidance mechanisms can be introduced, while targeting creativity stimulation. Moreover, we want to further extend this model with sentiment analysis features derived from local contexts captured by BERT, thus further enriching the analysis with the identification of convergent and divergent points of view.

List of Publications

ISI Proceedings

Top Ranked Conferences (category A in CORE Ranking Portal)

1. Ruseti, S., Dascalu, M.-D., Corlatescu, D.-G., Dascalu, M., Trausan-Matu, S., & McNamara, D. S. (in press). *Exploring Dialogism using Language Models*. In 22nd Int. Conf. on Artificial Intelligence in Education (AIED 2021). Utrech, Netherlands (Online): Springer.
2. Dascalu, M.-D., Ruseti, S., Dascalu, M., McNamara, D. S., & Trausan-Matu, S. (2020). *Multi-Document Cohesion Network Analysis: Visualizing Intratextual and Intertextual Links*. Paper presented at the 21st Int. Conf. on Artificial Intelligence in Education (AIED 2020), Online: Springer.
3. Dascalu, M.-D., Ruseti, S., Carabas, M., Dascalu, M., Trausan-Matu, S., & McNamara, D. S. (2020). *Cohesion Network Analysis: Predicting Course Grades and Generating Sociograms for a Romanian Moodle Course*. Paper presented at the 16h Int. Conf. on Intelligent Tutoring Systems (ITS 2020), Online: Springer. (Best Full Paper Award)
4. Samoilescu, R.-F., Dascalu, M., Sirbu, M.-D., Trausan-Matu, S., & Crossley, S. A. (2019). *Modeling Collaboration in Online Conversations using Time Series Analysis and Dialogism*. In 20th Int. Conf. on Artificial Intelligence in Education (AIED 2019) (pp. 458–468). Chicago, IL: Springer.
5. Crossley, S. A., Sirbu, M.-D., Dascalu, M., Barnes, T., Lynch, C. F., & McNamara, D. S. (2018). *Modeling Math Success using Cohesion Network Analysis*. In C. P. Rosé, R. Martínez-Maldonado, U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren & B. d. Boulay (Eds.), 19th Int. Conf. on Artificial Intelligence in Education (AIED 2018), Part II (pp. 63–67). London, UK: Springer.
6. Sirbu, M.-D., Dascalu, M., Crossley, S. A., McNamara, D. S., Barnes, T., Lynch, C. F., & Trausan-Matu, S. (2018). *Exploring Online Course Sociograms using Cohesion Network Analysis*. In C. P. Rosé, R. Martínez-Maldonado, U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren & B. d. Boulay (Eds.), 19th Int. Conf. on Artificial Intelligence in Education (AIED 2018), Part II (pp. 337–342). London, UK: Springer.

Other ISI Conferences

1. Dascalu, M.-D., Paraschiv, I. C., Nicula, B., Dascalu, M., Trausan-Matu, S., & Nuta, A. C. (2019). *Intelligent Platform for the Analysis of Drug Leaflets Using NLP Techniques*. In 18th Int.

Conf. on Networking in Education and Research (RoEduNet) (pp. 1–6). Galati, Romania: IEEE.

2. Ruseti, S., Sirbu, M.-D., Calin, M. A., Dascalu, M., Trausan-Matu, S., & Militaru, G. (2019). *Comprehensive Exploration of Game Reviews Extraction and Opinion Mining using NLP Techniques*. In 5th Int. Congress on Information and Communication Technology (pp. 323–331). London, UK: Springer.
3. Ene, O.-G., Sirbu, M.-D., Dascalu, M., Trausan-Matu, S., & Nuta, A. C. (2019). *PIAM – Intelligent Platform for Retrieving Relevant Information on Drugs Marketed in Romania*. In 4th Int. Workshop on Design and Spontaneity in Computer-Supported Collaborative Learning (DS-CSCL-2019), in conjunction with the 22nd Int. Conf. on Control Systems and Computer Science (CSCS22) (pp. 420–425). Bucharest, Romania: IEEE.
4. Stamati, D., Sirbu, M.-D., Dascalu, M., & Trausan-Matu, S. (2018). *Exploring General Morphological Analysis and Providing Personalized Recommendations to Stimulate Creativity with ReaderBench*. In M. Chang, E. Popescu, Kinshuk, N.-S. Chen, M. Jemni, R. Huang & J. M. Spector (Eds.), Int. Conf. on Smart Learning Environments (ICSLE 2018) (pp. 41–50). Beijing, China: Springer.
5. Sirbu, M.-D., Botarleanu, R., Dascalu, M., Crossley, S. A., & Trausan-Matu, S. (2018). *ReadME – Enhancing Automated Writing Evaluation*. In 18th Int. Conf. on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA 2018) (pp. 281–285). Varna, Bulgaria: Springer.
6. Dascalu, M., Sirbu, M.-D., Gutu-Robu, G., Ruseti, S., Crossley, S. A., & Trausan-Matu, S. (2018). *Cohesion-Centered Analysis of Sociograms for Online Communities and Courses using ReaderBench*. In 13th European Conference on Technology Enhanced Learning (EC-TEL 2018) (pp. 622–626). Leeds, UK: Springer.
7. Sirbu, M.-D., Panaite, M., Secui, A., Dascalu, M., Nistor, N., & Trausan-Matu, S. (2017). *ReaderBench: Building Comprehensive Sociograms of Online Communities*. In 19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2017) (pp. 225–231). Timisoara, Romania: IEEE.
8. Sirbu, M.-D., Secui, A., Dascalu, M., Crossley, S. A., Ruseti, S., & Trausan-Matu, S. (2016). *Extracting Gamers' Opinions from Reviews*. In 18th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2016) (pp. 227–232). Timisoara, Romania: IEEE.
9. Secui, A., Sirbu, M.-D., Dascalu, M., Crossley, S. A., Ruseti, S., & Trausan-Matu, S. (2016). *Expressing Sentiments in Game Reviews*. In 17th Int. Conf. on Artificial Intelligence:

Methodology, Systems, and Applications (AIMSA 2016) (pp. 352–355). Varna, Bulgaria: Springer.

10. Sirbu, M.-D., Dascalu, M., Gifu, D., Cotet, T.-M., Tosca, A., & Trausan-Matu, S. (2018). *ReadME – Improving Writing Skills in Romanian Language*. In V. Pais, D. Gifu, D. Trandabat, D. Cristea & D. Tufis (Eds.), 13th Int. Conf. on Linguistic Resources and Tools for Processing Romanian Language (ConsILR 2018) (pp. 135–145). Iasi, Romania.

BDI Proceedings

Top Ranked Conferences (category A in CORE Ranking Portal)

1. Sirbu, M.-D., Dascalu, M., Crossley, S. A., McNamara, D. S., & Trausan-Matu, S. (2019). *Longitudinal Analysis of Participation in Online Courses Powered by Cohesion Network Analysis*. In 13th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 2019) (pp. 640–643). Lyon, France: ISLS.

Other BDI Conferences

1. Fetoiu, C.-E., Dascalu, M.-D., Calin, M. A., Dascalu, M., Trausan-Matu, S., & Militaru, G. (2020). *Cohesion Network Analysis for Predicting User Ranks in Reddit Communities*. In 5th Int. Conf. on Smart Learning Ecosystems and Regional Development (SLERD 2020) (pp. 173–185). Online: Springer.
2. Paraschiv, I. C., Sirbu, M.-D., Nicula, B., Dascalu, M., Trausan-Matu, S., & Nuta, A. C. (2019). *Designing an Intelligent Platform for Drugs Administration*. In International Conference on Human-Computer Interaction (RoCHI2019) (pp. 53–59). Bucharest, Romania: MatrixRom.
3. Nicula, B., Dascalu, M., Sirbu, M.-D., Trausan-Matu, S., & Nuta, A. C. (2019). *Building a Comprehensive Romanian Knowledge Base for Drug Administration*. In G. Angelova, R. Mitkov, I. Nikolova & I. Temnikova (Eds.), Int. Conf. on Recent Advances in Natural Language Processing (RANLP 2019) (pp. 829–836). Varna, Bulgaria.
4. Florea, A.-M., Dascalu, M., Sirbu, M.-D., & Trausan-Matu, S. (2019). *Improving Writing for Romanian Language*. In 4th Int. Conf. on Smart Learning Ecosystems and Regional Development (SLERD 2019) (pp. 131–141). Rome, Italy: Springer.
5. Botarleanu, R.-M., Dascalu, M., Sirbu, M.-D., Crossley, S. A., & Trausan-Matu, S. (2018). *ReadME – Generating Personalized Feedback for Essay Writing using the ReaderBench Framework*. In H. Knoche, E. Popescu & A. Cartelli (Eds.), 3rd Int. Conf. on Smart Learning Ecosystems and Regional Development (SLERD 2018) (pp. 133–145). Aalborg, Denmark: Springer.
6. Dascalu, M.-D., Ruseti, S., Dascalu, M., McNamara, D. S., & Trausan-Matu, S. (in press). *Dialogism Meets Language Models for Evaluating Involvement in CSCL Conversations*. In 6th Int.

Conf. on Smart Learning Ecosystems and Regional Development (SLERD 2021) Bucharest, Romania (Online).

Journals

ISI Journals

1. Dascalu, M.-D., Ruseti, S., Dascalu, M., McNamara, D. S., Carabas, M., Rebedea, T., & Trausan-Matu, S. (2021). *Before and during COVID-19: A Cohesion Network Analysis of Students' Online Participation in Moodle Courses*. Computers in Human Behavior, 121 (**Q1 Journal**).
2. Dragomir, I. A., Dascalu, M.-D., Dascalu, M., Terian, S.-M., & Trausan-Matu, S. (2020). *Exploring Differences in NATO Discourses using the ReaderBench Framework*. Scientific Bulletin, University Politehnica of Bucharest, Series C, 82(1), 3–18.
3. Terian, S.-M., Cotet, T.-M., Sirbu, M.-D., Dascalu, M., & Trausan-Matu, S. (2019). *Discourses of Economic Crisis in Romanian Media: An Automated Analysis using the ReaderBench Framework*. Transylvanian Review, XXVIII (Supplement No. 1), 245–262.

Other Journals

1. Gutu-Robu, G., Sirbu, M.-D., Paraschiv, I. C., Dascalu, M., Dessus, P., & Trausan-Matu, S. (2018). *Liftoff - ReaderBench introduces new online functionalities*. Romanian Journal of Human - Computer Interaction, 11(1), 76–91.

Patent applications

1. Trausan-Matu, S., Dascalu, M., Dascalu, M.-D., Paraschiv, I. C., Nicula, B., Nuta, A. C., & Cioroboiu, G.-C. (2019). Romania Patent No. Cererea A/00877 / 09.12.2019. OSIM.

References

Achieve Virtual. (n.d.). Infographic: Evolution of Virtual Education. Retrieved September 1st, 2020, from https://achievevirtual.org/infographic-evolution-of-virtual-education/

Arif, T. (2015). The mathematics of social network analysis: metrics for academic social networks. *International Journal of Computer Applications Technology and Research, 4*(12), 889-893.

Attali, Y., & Burstein, J. (2004). Automated essay scoring with e-rater V.2.0. In *Annual Meeting of the International Association for Educational Assessment* (pp. 23). Philadelphia, PA: Association for Educational Assessment.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays* (C. Emerson & M. Holquist, Trans.). Austin, TX, USA and London, UK: The University of Texas Press.

Bakhtin, M. M. (1984). *Problems of Dostoevsky's poetics* (C. Emerson, Trans. C. Emerson Ed.). Minneapolis: University of Minnesota Press.

Bakhtin, M. M. (1986). *Speech genres and other late essays* (V. W. McGee, Trans.). Austin: University of Texas.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research, 3*(4-5), 993–1022.

Bolter, J. D. (1984). *Turing's man: Western culture in the computer age*. Chapel Hill, North Carolina, USA: UNC Press Books.

Bostock, M. (2021). Data-Driven Documents. Retrieved July, 22nd, 2021, from https://d3js.org/

Botarleanu, R.-M., Dascalu, M., Sirbu, M.-D., Crossley, S. A., & Trausan-Matu, S. (2018). ReadME – Generating Personalized Feedback for Essay Writing using the ReaderBench Framework. In H. Knoche, E. Popescu & A. Cartelli (Eds.), *3rd Int. Conf. on Smart Learning Ecosystems and Regional Development (SLERD 2018)* (pp. 133–145). Aalborg, Denmark: Springer.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems, 30*, 1–7.

Britannica, E. (2021). Britannica – University of Phoenix. Retrieved June, 9th, 2021, from https://www.britannica.com/topic/University-of-Phoenix

Cambridge Intelligence. (2021). Social Network Analysis. Retrieved May, 30th, 2021, from https://cambridge-intelligence.com/social-network-analysis/

Chowdhury, G. G. (2003). Natural language processing. *Annual Review of Information Science and Technology, 37*(1), 51-89.

Clark, K., Khandelwal, U., Levy, O., & Manning, C. D. (2019). What does bert look at? an analysis of bert's attention. *arXiv preprint arXiv:1906.04341*.

References

College, H. (2021). The Writing Process. Retrieved July, 14th, 2021, from http://www.hunter.cuny.edu/rwc/handouts/the-writing-process-1/invention/Five-Qualities-of-Good-Writing

Cress, U. (2013). Mass collaboration and learning. In R. Luckin, S. Puntambekar, P. Goodyear, B. Grabowski, J. Underwood & N. Winters (Eds.), *Handbook of design in educational technology* (pp. 416–424). New York: Routledge.

Crossley, S. A., Barnes, T., Lynch, C., & McNamara, D. S. (2017). Linking language to math success in a blended course. In *10th Int. Conf, on Educational Data Mining (EDM)* (pp. 180–185). Wuhan, China.

Crossley, S. A., Kyle, K., & Dascalu, M. (2019). The Tool for the Automatic Analysis of Cohesion 2.0: Integrating Semantic Similarity and Text Overlap. *Behavior research methods, 51*(1), 14–27.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). To aggregate or not? Linguistic features in automatic essay scoring and feedback systems. *Journal of Writing Assessment, 8*(1), 19.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2016). The tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior research methods, 48*(4), 1227–1237.

Crossley, S. A., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining Click-Stream Data with NLP Tools to Better Understand MOOC Completion. In *6th Int. Conf. on Learning Analytics & Knowledge (LAK '16)* (pp. 6–14). Edingurgh, UK: ACM.

Crossley, S. A., Sirbu, M.-D., Dascalu, M., Barnes, T., Lynch, C. F., & McNamara, D. S. (2018). Modeling Math Success using Cohesion Network Analysis. In C. P. Rosé, R. Martínez-Maldonado, U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren & B. d. Boulay (Eds.), *19th Int. Conf. on Artificial Intelligence in Education (AIED 2018), Part II* (pp. 63–67). London, UK: Springer.

Dale, E., & Chall, J. S. (1949). The concept of readability. *Elementary English, 26*(1), 19-26.

Das, K., Samanta, S., & Pal, M. (2018). Study on centrality measures in social networks: a survey. *Social network analysis and mining, 8*(1), 1-11.

Dascalu, M. (2014). *Analyzing discourse and text complexity for learning and collaborating, Studies in Computational Intelligence* (Vol. 534). Switzerland: Springer.

Dascalu, M., Crossley, S. A., McNamara, D. S., Dessus, P., & Trausan-Matu, S. (2018). Please ReaderBench this Text: A Multi-Dimensional Textual Complexity Assessment Framework. In S. Craig (Ed.), *Tutoring and Intelligent Tutoring Systems* (pp. 251–271). Hauppauge, NY, USA: Nova Science Publishers, Inc.

Dascalu, M., Dessus, P., Bianco, M., Trausan-Matu, S., & Nardy, A. (2014). Mining texts, learner productions and strategies with ReaderBench. In A. Peña-Ayala (Ed.), *Educational Data Mining: Applications and Trends* (pp. 345–377). Cham, Switzerland: Springer.

Dascalu, M., McNamara, D. S., Trausan-Matu, S., & Allen, L. K. (2018). Cohesion Network Analysis of CSCL Participation. *Behavior research methods, 50*(2), 604–619. doi: 10.3758/s13428-017-0888-4

Dascalu, M., Sirbu, M. D., Gutu-Robu, G., Ruseti, S., Crossley, S. A., & Trausan-Matu, S. (2018). Cohesion-Centered Analysis of Sociograms for Online Communities and Courses using ReaderBench. In *13th European Conference on Technology Enhanced Learning (EC-TEL 2018)* (pp. 622–626). Leeds, UK: Springer.

Dascalu, M., Trausan-Matu, S., & Dessus, P. (2013). Voices' inter-animation detection with ReaderBench – Modelling and assessing polyphony in CSCL chats as voice synergy. In *2nd Int. Workshop on Semantic and Collaborative Technologies for the Web, in conjunction with the 2nd Int. Conf. on Systems and Computer Science (ICSCS)* (pp. 280–285). Villeneuve d'Ascq, France: IEEE.

Dascalu, M., Trausan-Matu, S., McNamara, D. S., & Dessus, P. (2015). ReaderBench – Automated Evaluation of Collaboration based on Cohesion and Dialogism. *International Journal of Computer-Supported Collaborative Learning, 10*(4), 395–423.

Dascalu, M.-D., Paraschiv, I. C., Nicula, B., Dascalu, M., Trausan-Matu, S., & Nuta, A. C. (2019). Intelligent Platform for the Analysis of Drug Leaflets Using NLP Techniques. In *18th Int. Conf. on Networking in Education and Research (RoEduNet)* (pp. 1–6). Galati, Romania: IEEE.

Dascalu, M.-D., Ruseti, S., Carabas, M., Dascalu, M., Trausan-Matu, S., & McNamara, D. S. (2020). Cohesion Network Analysis: Predicting Course Grades and Generating Sociograms for a Romanian Moodle Course. In *16h Int. Conf. on Intelligent Tutoring Systems (ITS 2020)*. Online: Springer.

Dascalu, M.-D., Ruseti, S., Dascalu, M., McNamara, D. S., Carabas, M., Rebedea, T., & Trausan-Matu, S. (2021). Before and during COVID-19: A Cohesion Network Analysis of Students' Online Participation in Moodle Courses. *Computers in Human Behavior, 121*.

Dascalu, M.-D., Ruseti, S., Dascalu, M., McNamara, D. S., & Trausan-Matu, S. (2020). Multi-Document Cohesion Network Analysis: Visualizing Intratextual and Intertextual Links. In *21st Int. Conf. on Artificial Intelligence in Education (AIED 2020)*. Online: Springer.

Davies, M. (2010). The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English. *Literary and Linguistic Computing, 25*(4), 447–465.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NACL 2018)*. Minneapolis, Minnesota, USA: ACL.

Dragomir, I. A., Dascalu, M.-D., Dascalu, M., Terian, S.-M., & Trausan-Matu, S. (2020). Exploring Differences in NATO Discourses using the ReaderBench Framework. *Scientific Bulletin, University Politehnica of Bucharest, Series C, 82*(1), 3–18.

DuBay, W. H. (2004). *The Principles of Readability*. Costa Mesa, CA, USA: Impact Information.

References

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology, 38*(1), 188–230.

Ene, O.-G., Sirbu, M.-D., Dascalu, M., Trausan-Matu, S., & Nuta, A. C. (2019). PIAM – Intelligent Platform for Retrieving Relevant Information on Drugs Marketed in Romania. In *4th Int. Workshop on Design and Spontaneity in Computer-Supported Collaborative Learning (DS-CSCL-2019), in conjunction with the 22nd Int. Conf. on Control Systems and Computer Science (CSCS22)* (pp. 420–425). Bucharest, Romania: IEEE.

Explosion. (2016-2021). spaCy. Retrieved July, 27th, 2021, from https://spacy.io

Fetoiu, C.-E., Dascalu, M.-D., Calin, M. A., Dascalu, M., Trausan-Matu, S., & Militaru, G. (2020). Cohesion Network Analysis for Predicting User Ranks in Reddit Communities. In *5th Int. Conf. on Smart Learning Ecosystems and Regional Development (SLERD 2020)* (pp. 173–185). Bucharest, Romania (Online): Springer.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 35-41.

Gabriel, C., Hahne, C., Zimmermann, A., & Lenk, F. (2021). The Virtual Tutor: Tasks for conversational agents in Online Collaborative Learning Environments. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 104–113). Hawaii, USA: Online http://hdl.handle.net/10125/70623.

Galley, M., & McKeown, K. (2003). Improving word sense disambiguation in lexical chaining. In G. Gottlob & T. Walsh (Eds.), *18th International Joint Conference on Artificial Intelligence (IJCAI'03)* (pp. 1486–1488). Acapulco, Mexico: Morgan Kaufmann Publishers, Inc.

Google Inc. (2010-2021). Angular website. Retrieved July, 22nd, 2021, from https://angular.io

Gutu-Robu, G., Sirbu, M.-D., Paraschiv, I. C., Dascalu, M., Dessus, P., & Trausan-Matu, S. (2018). Liftoff - ReaderBench introduces new online functionalities. *Romanian Journal of Human - Computer Interaction, 11*(1), 76–91.

Hansen, D. L., Shneiderman, B., Smith, M. A., & Himelboim, I. (2011). Social network analysis: measuring, mapping, and modeling collections of connections. In B. Shneiderman, D. Hansen & M. A. Smith (Eds.), *Analyzing social media networks with NodeXL* (pp. 31–51). Amsterdam, Netherlands: Elsevier.

Hiltz, S. R., & Turoff, M. (2005). Education goes digital: The evolution of online learning and the revolution in higher education. *Communications of the ACM, 48*(10), 59–64.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735–1780.

Holmes, D. I. (1998). The evolution of stylometry in humanities scholarship. *Literary and Linguistic Computing, 13*(3), 111-117.

John Snow Labs Inc. (2021). Spark NLP. Retrieved July, 5th, 2021, from https://nlp.johnsnowlabs.com

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260.

Jurafsky, D., & Martin, J. H. (2008). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics* (2nd ed.). London: Pearson Prentice Hall.

Kaggle Inc. (2021). Automated Student Assessment Prize. Retrieved July, 5th, 2021, from https://www.kaggle.com/c/asap-aes/overview/description

Klecka, W. R. (1980). Discriminant analysis *Quantitative Applications in the Social Sciences Series* (Vol. 19). Thousand Oaks, CA: Sage Publications.

Knoke, D., & Yang, S. (2019). *Social network analysis*: Sage Publications.

Koschmann, T. (1996). Paradigm shifts and instructional technology: an introduction. In T. Koschmann (Ed.), *CSCL, Theory and Practice of an Emerging Paradigm*. Mahwah, New Jersey, USA: Routledge.

Koschmann, T. (1999). Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In C. M. Hoadley & J. Roschelle (Eds.), *Int. Conf. on Computer Support for Collaborative Learning (CSCL'99)* (pp. 308–313). Palo Alto: ISLS.

Kyle, K. (2016). Measuring syntactic development in L2 writing: Fine grained indices of syntactic complexity and usage-based indices of syntactic sophistication.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*, 757–786. doi: 10.1002/tesq.194

Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the automatic analysis of lexical sophistication (TAALES): version 2.0. *Behavior research methods, 50*(3), 1030–1046. doi: 10.3758/s13428-017-0924-4

Kyle, K., Crossley, S. A., & Jarvis, S. (2020). Assessing the Validity of Lexical Diversity Indices Using Direct Judgements. *Language Assessment Quarterly*, 1-17.

Landauer, T. K., & Dumais, S. (2008). Latent semantic analysis. *Scholarpedia, 3*(11), 4356.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: the Latent Semantic Analysis theory of acquisition, induction and representation of knowledge. *Psychological Review, 104*(2), 211–240.

LGPLv2.1, G. (2009-2021). Gensim. Retrieved July, 5th, 2021, from https://pypi.org/project/gensim/

Light, R. J. (2004). *Making the most of college: Students speak their minds*: Harvard University Press.

Linell, P. (2009). *Rethinking language, mind, and world dialogically: Interactional and contextual theories of human sense-making*. Information Age Publishing: Charlotte, NC.

Lipponen, L. (2002). Exploring foundations for computer-supported collaborative learning.

Liu, Z., Lin, Y., & Sun, M. (2020). Word Representation *Representation Learning for Natural Language Processing* (pp. 13-41): Springer.

Marková, I., Linell, P., Grossen, M., & Salazar Orvig, A. (2007). *Dialogue in focus groups: Exploring socially shared knowledge*. London, UK: Equinox.

McAuley, A., Stewart, B., Siemens, G., & Cormier, D. (2010). The MOOC model for digital practice.

References

McNamara, D. S. (2004). SERT: Self-Explanation Reading Training. *Discourse Processes, 38*, 1–30.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). The linguistic features of quality writing. *Written Communication, 27*(1), 57–86.

McNamara, D. S., Crossley, S. A., & Roscoe, R. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior research methods, 45*(2), 499-515.

McNamara, D. S., Crossley, S. A., Roscoe, R., Allen, L. K., & Dai, J. (2015). A hierarchical classification approach to automated essay scoring. *Assessing Writing, 23*, 35–59.

McNamara, D. S., Graesser, A. C., & McCarthy, P. M. (2014). Coh-Metrix. Retrieved May, 26th, 2021, from http://cohmetrix.com

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representation in Vector Space. In *Workshop at ICLR*. Scottsdale, AZ.

Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM, 38*(11), 39–41.

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of machine learning*: MIT press.

Moodle. (2021). Moodle website. Retrieved May, 28th, 2021, from https://moodle.org

Moore, J. L., Dickson-Deane, C., & Galyen, K. (2011). e-Learning, online learning, and distance learning environments: Are they the same? *The Internet and Higher Education, 14*(2), 129–135.

Moravcsik, J. E., & Kintsch, W. (1993). Writing quality, reading skills, and domain knowledge as factors in text comprehension. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 47*(2), 360.

Mukherjee, P., Leroy, G., & Kauchak, D. (2018). Using Lexical Chains to Identify Text Difficulty: A Corpus Statistics and Classification Study. *IEEE journal of biomedical and health informatics, 23*(5), 2164-2173.

Nicula, B., Dascalu, M., Sirbu, M.-D., Trausan-Matu, S., & Nuta, A. C. (2019). Building a Comprehensive Romanian Knowledge Base for Drug Administration. In G. Angelova, R. Mitkov, I. Nikolova & I. Temnikova (Eds.), *Int. Conf. on Recent Advances in Natural Language Processing (RANLP 2019)* (pp. 829–836). Varna, Bulgaria.

Nistor, N., Dascalu, M., Serafin, Y., & Trausan-Matu, S. (2018). Automated Dialog Analysis to Predict Blogger Community Response to Newcomer Inquiries. *Computers in Human Behavior, 89*, 349–354. doi: https://doi.org/10.1016/j.chb.2018.08.034

Nistor, N., Dascalu, M., Tarnai, C., & Trausan-Matu, S. (2020). Predicting newcomer integration in online learning communities: Automated dialog assessment in blogger communities. *Computers in Human Behavior, 105*.

O'Rourke, S. T., & Calvo, R. A. (2009). Analysing semantic flow in academic writing. In V. Dimitrova, R. Mizoguchi, B. du Boulay & A. C. Graesser (Eds.), *Artificial Intelligence in Education. Building learning* *systems that care: From knowledge representation to affective modelling (AIED2009)* (pp. 173–180). Amsterdam, The Netherlands: IOS Press.

O'Reilly, T., & McNamara, D. S. (2007). Reversing the reverse cohesion effect: Good texts can be better for strategic, high-knowledge readers. *Discourse Processes, 43*(2), 121–152.

Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web: Stanford InfoLab.

Paraschiv, I. C., Sirbu, M.-D., Nicula, B., Dascalu, M., Trausan-Matu, S., & Nuta, A. C. (2019). Designing an Intelligent Platform for Drugs Administration. In *International Conference on Human-Computer Interaction (RoCHI2019)* (pp. 53–59). Bucharest, Romania: MatrixRom.

Rasa Technologies Inc. (2021). Rasa website. Retrieved July, 5th, 2021, from https://rasa.com

Rice, J. (2018). What Makes Hemingway Hemingway? *The LitCharts Blog, 3*.

Roscoe, R. D., Varner, L. K., Crossley, S. A., & McNamara, D. S. (2013). Developing pedagogically-guided algorithms for intelligent writing feedback. *International Journal of Learning Technology 25, 8*(4), 362–381.

Scott, J. (1988). Social network analysis. *Sociology, 22*(1), 109-127.

Scott, J. (2017). *Social Network Analysis*. London, UK: Sage.

Shaw, M. E. (1954). Group structure and the behavior of individuals in small groups. *The Journal of psychology, 38*(1), 139-149.

Simonson, M. (2007). Course management systems. *Quarterly review of distance education, 8*(1), 7-9.

Sirbu, M.-D., Botarleanu, R., Dascalu, M., Crossley, S. A., & Trausan-Matu, S. (2018). ReadME – Enhancing Automated Writing Evaluation. In *18th Int. Conf. on Artificial Intelligence: Methodology, Systems, and Applications (AIMSA 2018)* (pp. 281–285). Varna, Bulgaria: Springer.

Sirbu, M.-D., Dascalu, M., Crossley, S. A., McNamara, D. S., Barnes, T., Lynch, C. F., & Trausan-Matu, S. (2018a). Exploring Online Course Sociograms using Cohesion Network Analysis. In C. P. Rosé, R. Martínez-Maldonado, U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren & B. d. Boulay (Eds.), *19th Int. Conf. on Artificial Intelligence in Education (AIED 2018), Part II* (pp. 337–342). London, UK: Springer.

Sirbu, M.-D., Dascalu, M., Crossley, S. A., McNamara, D. S., & Trausan-Matu, S. (2019). Longitudinal Analysis of Participation in Online Courses Powered by Cohesion Network Analysis. In *13th Int. Conf. on Computer-Supported Collaborative Learning (CSCL 2019)* (pp. 640–643). Lyon, France: ISLS.

Sirbu, M.-D., Dascalu, M., Gifu, D., Cotet, T.-M., Tosca, A., & Trausan-Matu, S. (2018b). ReadME – Improving Writing Skills in Romanian Language. In V. Pais, D. Gifu, D. Trandabat, D. Cristea & D. Tufis (Eds.), *13th Int. Conf. on Linguistic Resources and Tools for Processing Romanian Language (ConsILR 2018)* (pp. 135–145). Iasi, Romania.

References

Sirbu, M.-D., Panaite, M., Secui, A., Dascalu, M., Nistor, N., & Trausan-Matu, S. (2017). ReaderBench: Building Comprehensive Sociograms of Online Communities. In *19th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC 2017)* (pp. 225–231). Timisoara, Romania: IEEE.

Somasundaran, S., Burstein, J., & Chodorow, M. (2014). Lexical chaining for measuring discourse coherence quality in test-taker essays. In *Proceedings of COLING 2014, the 25th International conference on computational linguistics: Technical papers* (pp. 950-961).

Stahl, G. (2006). *Group cognition. Computer support for building collaborative knowledge.* Cambridge, MA: MIT Press.

Stamatatos, E. (2009). A survey of modern authorship attribution methods. *Journal of the American Society for Information Science and Technology, 60*(3), 538-556.

Stamati, D., Dascalu, M., & Trausan-Matu, S. (2015). Creativity stimulation in chat conversations through morphological analysis. *University Politehnica of Bucharest Scientific Bulletin Series C-Electrical Engineering and Computer Science, 77*(4), 17–30.

Stanford NLP Group. (2021). Stanford CoreNLP. Retrieved July, 5th, 2021, from https://stanfordnlp.github.io/CoreNLP/

Terian, S.-M., Cotet, T.-M., Sirbu, M.-D., Dascalu, M., & Trausan-Matu, S. (2019). Discourses of Economic Crisis in Romanian Media: An Automated Analysis using the ReaderBench Framework. *Transylvanian Review, XXVIII*(Supplement No. 1), 245–262.

The Allen Institute for Artificial Intelligence. (2021). Allen NLP. Retrieved July, 5th, 2021, from https://allennlp.org

Trausan-Matu, S. (2010). Automatic Support for the Analysis of Online Collaborative Learning Chat Conversations. In P. M. Tsang, S. K. S. Cheung, V. S. K. Lee & R. Huang (Eds.), *3rd Int. Conf. on Hybrid Learning* (pp. 383–394). Beijing, China: Springer.

Trausan-Matu, S. (2020). The Polyphonic Model of Collaborative Learning. In N. Mercer, R. Wegerif & L. Major (Eds.), *The Routledge International Handbook of Research on Dialogic Education* (pp. 454–468). Abingdon, UK: Routledge.

Trausan-Matu, S., Dascalu, M., Rebedea, T., & Gartner, A. (2010). Corpus de conversatii multi-participant si editor pentru adnotarea lui. *Revista Romana de Interactiune Om-Calculator, 3*(1), 53–64.

Trausan-Matu, S., & Rebedea, T. (2009). Polyphonic inter-animation of voices in VMT. In G. Stahl (Ed.), *Studying Virtual Math Teams* (pp. 451–473). Boston, MA: Springer.

Trausan-Matu, S., Stahl, G., & Sarmiento, J. (2007). Supporting polyphonic collaborative learning. *E-service Journal, Indiana University Press, 6*(1), 58–74.

Tu, C.-H. (2002). The measurement of social presence in an online learning environment. *International Journal on E-learning, 1*(2), 34–45.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *31st Conf. on Neural Information Processing Systems (NIPS 2017)* (pp. 5998–6008). Long Beach, CA, USA.

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. *Language Teaching Research, 10*, 1–24.

Watson, W., & Watson, S. L. (2007). An argument for clarity: What are learning management systems, what are they not, and what should they become. *TechTrends, 51*(2), 28-34.

Weidlich, J., & Bastiaens, T. J. (2019). Designing sociable online learning environments and enhancing social presence: An affordance enrichment approach. *Computers & Education, 142*, 16.

Witte, S. P., & Faigley, L. (1981). Coherence, cohesion, and writing quality. *College composition and communication, 32*(2), 189-204.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The educator's word frequency guide*. Brewster, NY: Touchstone Applied Science Associates, Inc.