From Traditional Linguistics to Computational Linguistics. The Relevance of Digital Corpuses in Education

Marius OPINCARIU

Universitatea „Lucian Blaga" din Sibiu, Facultatea de Litere și Arte
Lucian Blaga University of Sibiu, Faculty of Letters and Arts
Personal e-mail: opincariugroup@gmail.com

From Traditional Linguistics to Computational Linguistics.
The Relevance of Digital Corpuses in Education

Recent advances in natural language processing architectures open new opportunities for enhanced educational designs. From proto-linguistics to natural language processing, the new era of internet-linguistics facilitates significant progress in the fields of computational linguistics and annotated digital corpora. With the help of quantitative linguistics, digital text compositions gain increased relevance in contemporary educational discourse analysis via computational semantics and word sense disambiguation. Digital linguistics may provide key performance indicators to fields such as higher education where written contributions are of critical importance.

Keywords: responsive e-learning, educational data mining, syntactic maps, cognitive curriculum calibration, affective curriculum calibration.

1. The State of Innovation in Education

Could one argue that the lack of disruptive actors in the field of education will leave teaching, learning and policy design largely unaffected? In a humble attempt to answer this question, the present work aims to offer a brief refresh of the existing state of education as well as to point at recent technological developments primed to enhance curriculum design.

A growing consensus points that current educational silo data management models outstrip educational infrastructure of the flexibility and adaptive reframing capabilities required to address the rapidly changing economic scenarios. Slowly formulated macro-policies lead to order of magnitude gaps between educational output and workforce input. This is why, in face of the new challenges, recurrent patterns increasingly draw the attention and capture the interest of experts in the field of education research. In the context of technologically enhanced models of communication, a fair amount of discussion resurfaces around the role of artificial intelligence in education. It may be related to the fact that AI is expected to have a major impact on the near future world of work, which requires a more technologically reformulated academic output. If in the past the fundamental blocks of curriculum architecture were designed to produce an academic output addressing simpler levels of automation, the current educational institutions experience the pressure of addressing emerging problems inexistent in the industrial age. As some of the approaches are gradually becoming outdated,[1] the research literature points to a paradigm which integrates critical edutech concepts with market forces markers. It has therefore become customary to see that when education and data driven governance are incorporated in the same conceptual design, they are either linked to workforce development, educational data mining and learning analytics (Cope & Kalantzis, 2016; Gagliardi, Parnell & Carpenter-Hubin, 2018; Williamson, 2017), or to the way in which disruptive innovation has challenged the orthodoxy of the classroom (Christensen, Horn & Johnson, 2008).[2] These innovative instruments are anticipated to field more alignment points between education output and market requirements. The proposed architectures intend to reveal more detailed information about the misaligned data points between the two verticals and provide more complex critical parameter performance visualisation maps. Such a vision is seeded in the grounds that integrate platformized with traditional learning, facilitating the discovery of unobservable patterns in data flows.

There is an estimated number of three levels of social practices where automation and smart architectures come in contact with educational design, namely, operations, acts and activities. The current rhetoric centers around operations, where elearning platforms facilitate the integration of algorithmic architectures. This level responds to the question of "how." Precisely, it implements acts in concrete settings. As there are numerous ways of assessing either student skills, assignment types or delivery methods, the "operations" level is the milieu where technological functioning as a tool meets the human behaviour in its routine, repetitive and habitual states.[3] It is at this level that AI augments and complements social practices, increasing the efficiency and effectiveness of predefined processes. As it progresses through the next level of "acts," it replaces, substitutes, and automates acts that were previously done by humans. When it reaches the highest level of social practice at the level of "activity," it transforms the system of motives, making current activities and specializations redundant and obsolete.[4]

While the available and applicable tech is positioned closer to the middle level of acts, an optimistic level of readiness for AI implementation in the field of education, is at the level of "operations." Here, vocational education has traditionally focused on technical and routine skills by teaching students how to use tools and domain-specific knowledge. The recent trends however, call for more competence-based educational designs which focus on problem solving, critical thinking, decision-making and analytical skills, all reflective of the cognitive level. If the cognitive design corresponds to acts, entrepreneurial and innovation competences are likely to be linked with the highest echelon of "activities." This is the space where social and cultural change occurs, where frameworks for developing the competencies and 21st century skills are incubated and developed. With the gradual prospect of responsive architectures impacting education, the upcoming landscape points to a major paradigm shift of the role of technology in education. If the industrial age was focused more on tools for automating and supporting manufacturing, the redesigned vision incorporates technologies aimed at social change.[5] However, in spite of a push for a technologically enhanced social change, the current neural AI and machine learning systems only address the bottom level of the three-level hierarchy. Namely, only the tasks that require habit formation and reflex reaction are well suited for supervised learning models.[6]

2. The Impact of AI on Teaching, Learning and Policy Design

2.1. The Impact on Teaching and Learning

Among the important questions becoming relevant in the upcoming educational decade however, one is how to make technology more pedagogically useful. As digital environments gradually fluidize the boundaries between academic disciplines, there is a growing argument for an interdisciplinary unified vision. To stay relevant and deliver an economically viable product, the educational infrastructure is to become fully integrated as a digital vertical along with other economic sectors which are gradually reinvented as smart mobility, smart care, smart retail, smart energy, smart homes, smart building management, smart manufacturing, IOT integration, etc. Within the boundaries of a smart society, education will eventually morph into smart education. Beyond the pure purpose of advancing knowledge, the vertical integration aims primarily at channeling academic research into industrial output,[7] or at the least, maintaining a high degree of correlation between the two. The *smart* [8] concept penetrates the traditional hierarchy very early on, and surfaces both at the primary and secondary levels of education.

> "Which vocations and occupations will become obsolete in the near future? What are the 21st century skills in a world where AI is widely used? How should AI be incorporated in the K-12 curriculum? How will AI change teaching? Should real-time monitoring of student emotions be allowed in classrooms? Can AI fairly assess students? Do we need fewer classrooms because of AI? Does AI reduce the impact of dyslexia, dyscalculia, or other learning difficulties?"[9]

Such complex questions require answers based on deep learning designs. It is not a secret anymore that AI systems excell at combining evidence from complex data sources that result in real time pattern recognition processing. Examples range from student homework that can be verified by an AI system possessing data on both the individual student history and peer responses, to highly summative or formative assessments. Of particular importance are the scientific disciplines where thresholds can be easily benchmarked. In the cases of mathematics, chemistry, physics etc., high-stakes testing functions can be automated. If the science tasks require simpler operating architectures, behavior and cognition processing raises the bar in terms of architecture complexity. For such an assignment, mathematical psychology models can be deployed to address the perceptual, thought, cognitive and motor processes, with rules that connect quantifiable stimulus characteristics with quantifiable behavior.[10] These systems can be easily integrated with elearning platforms which are natively designed for human-computer interaction. AI is being increasingly employed for student attention diagnostics, emotion, and conversation dynamics in computer-supported learning environments. Such systems can also prove useful in course development and management where they can generate and suggest collaborative learning patterns and optimal group formation by identifying and classifying group cohesion and possible student drop-out parameters (Nkambou et al. 2018; Rosé et al. 2018).[11] By understanding group patterns of interaction, neural networks are able to correlate them with pedagogically relevant classifications, making tutor interventions and guidance more precise. AI can apply diagnostic data evaluations and reflect back to the students their data based metacognitive approaches, suggesting individual areas of potential improvement. Neural networks are posed therefore to deliver important pedagogical insights by enhancing educational data mining designs with learning analytics and diagnostics.[12] The introduction of natural language processing in the educational designs adds unprecedented levels of access to data exploration, visualisation and modeling. The added function of the new enhancement, improves on the unidirectional educational formula by integrating metrics on both tutor input and student interactions.

2.2 The impact on Educational Policy Design

If big data, artificial intelligence, machine learning and predictive analytics continue to shape the industry verticals, the digital educational data will gradually be transformed by AI in EI. This facilitates the gradual integration of technology in educational governance enabling the growth of educational intelligence (EI) applications.

> "Defined by its core features such as volume, velocity and veracity, exhaustiveness, flexibility and scalability (Kitchin 2014), considers Big Data as a disruptive innovation powerful enough that "a data revolution is underway and has far-reaching consequences, to how knowledge is produced, business conducted, and governance enacted."[13]

In a government conference aimed at fostering innovation, SoftBank's CEO Masayoshi Son argued that Japan should make AI a mandatory subject for college entrance exams, particularly because japanese students "don't study if they are not asked... let's put it as mandatory, then Japanese students will catch up."[14] The strategic importance of education is therefore not new, as the terminology was first coined as *educational intelligence* at the beginning of the 20th century when Sir Michael Sadler, a brilliant educationist, proposed elevating education at the level of importance of other critical national security sectors. Instead of the UK Office of Special Inquiries and Reports, responsible for comparative education analysis, he proposed:

> "[...] the Intelligence Department of the Board of Education. It would also more closely conform to the nomenclature adopted for the Intelligence Division of the Admiralty, War Office and the Board of Trade (Sadler, as cited in Sislian, 2004, p. 8)."[15]

Conceptualizing educational intelligence as a strategic asset equal in importance to military and economic intelligence, has securely kept the United Kingdom's educational infrastructure on the world class educational stage for the last decades. While at that time printed documents only allowed for limited educational data modeling, at the present time, both the vast digital data quantities and the highly diversified modern instruments offer education an opportunity for exploration on a strategic level. As of 2020, big data can offer flexible algorithmization and computerisation opportunities for educational infrastructure and policy design.

Smart societies require adaptive educational infrastructures able to rapidly deploy economically viable workforce. The necessity of standardizing the teaching and learning outcomes in conformity with the OECD PISA's 21st century skills, therefore becomes a growing imperative. Important efforts made in this direction signal the relevance of educational data for leveraging both intranational and international outcomes:

> "[...] analogous to Revel's (2010) concept of economic intelligence, educational intelligence represents a prized asset in this econosphere, as national educational systems and international organizations, such as the OECD or the EU seek to: (i) master the utilisation of Big Data for forecasting and planning educational reforms; (ii) determine the financing of education based on performance, accountability and benchmarking; and (iii) influence educational policy to effect adjustments or improvements in educational systems."[16]

If in the not so distant past, educational intelligence was limited to determining the competitiveness of educational systems between nation-states, during the fourth industrial revolution however, EI could be instrumental in delivering a unified european vision.[17] Adopting a data driven governance strategy at the european level, would facilitate the evolution of comprehensive multi-state data compilation and value extraction designs addressing both the historical and contemporary data. As the current european demographics, reformulate the national as regional and the interstate as intranational, fresh approaches may become relevant.

> "[...] the world is witnessing the steady and gradual transition from a knowledge economy or society to an educational intelligent economy premised on the exponential production of digital data, to measure, analyze and predict educational performance in comparative perspective. Furthermore, the digitization and datafication of educational output in the "data-driven, algorithm-mediated economy of the 21st century" (Economist, 2019, p. 1) have intensified the processing and analysis of data leading to the emergence of a form of digital education governance through massive flows of "Big-Data" (Williamson, 2017). Given its perceived value and potential to engender unforeseen challenges, this type of data may be construed as educational intelligence, to be exchanged, exploited, and leveraged for a multitude of purposes in the global educational markets and worlds of policy making. We posit that at the heart of an intelligent economy is educational intelligence which ecompasses both individual and system-level processes. This has the real potential of unleashing the creative capacity of educational systems to find innovative solutions in harnessing the learning required to manage and steer data integration at the intersection of Big Data, cloud computing, social media, mobile and automation technologies, and scientific discoveries that continuously reshape the way we live, work and learn."[18]

EI therefore is a concept that supersedes the limited understanding of siloed techno-statistical relevance, and is to be regarded as a long term strategic asset. In such a scenario, smart education would rise at the importance level of national strategic sectors, playing a determinant role in shaping economic interests.

Whether supervised or unsupervised, artificial intelligence architectures have grown to possess important algorithmic capabilities fit for enhancing rapid structural changes. AI has evolved beyond the basic tool employed for simple computations into a cross-functional instrument designed to equally effectively supervise and automate industrial or educational processes. In spite of a positively evolving legal framework, costs, systemic resistance to innovation and change as well as lack of qualified personnel makes implementation slow, sending a risky signal of failed acceptance. Recent studies however, point at more ingrained obstacles prohibiting the adoption of EI designs. One of them is the research community's preconceived reticence and bias towards AI framed research evidence.

> "[...] when educational research evidence is framed within AI research, it is considered as less credible in comparison to when it is framed instead within neuroscience or educational psychology [...] the results of the study indicate that the general public perceives AI to be: less helpful in assisting us to understand how children learn, lacking in adherence to scientific methods, and to be less prestigious compared to neuroscience and educational psychology [...] with a prevailing public image of AI being less scientifically robust and less prestigious than educational psychology and neuroscience."[19]

As long as AI is perceived as an existential threat and not as an instrument for augmentation and growth, systemic resistance is likely to prevail. As AI will be increasingly deployed to automate manufacturing processes, pressure to reinvent current educational institutions may become more stringent. It is, for example, possible that formal education will play a diminishing role in creating job-related competencies. This may suggest that the future role of education will increasingly be in competency development that is cognitive, affective and conative, known for supporting human development and be less invested in competence based basic skill development. A growing consensus seems to point that as educational systems remain adapted to the requirements of the industrial age, AI could make some functions of traditional education obsolete. While emphasizing others, it may also enable new ways of teaching and learning.[20]

While not discounting the cost and discomfort of change, the actionable technology behind the fancy terminology may provide more opportunities for learning and growth rather than for confusion and loss. Uncertainty around terminology and the general misunderstanding of the concepts are inducive of an unfounded fear that implemented AI systems will lead to widespread destruction. However, an alternating sum of benefits and possible challenges may offer a picture of the 21st century educational development.

> "The use of AI in education may generate insights on how learning happens, and it can change the way learning is assessed. It may re-organize classrooms or make them obsolete, it can increase the efficiency of teaching, or it may force students to adapt to the requirements of technology."[21]

In significant ways, most of the research domains are being increasingly impacted by discoveries in the fields of brain imaging and processing power. Modern policy design must therefore provide a balanced framework which prepares the society for a transparent and natural adoption of smart architectures.

> "For the effective adoption of AI, it is important to create a shared understanding between the key stakeholders of AI Technologies including public, educators, and academia (Cukurova et al. 2019)[22] [...] As a result, AI will probably have its biggest impact when it is used to augment human cognition, and in supporting human learning and knowing."[23]

Morphing from a directive to a supportive design, smart learning designs will gradually impersonate the role of a coach and mediator. The upcoming educational designs can only come to life by successfully marrying the advanced computational processing power with data mining and visualisation. Such instruments can be natively integrated into platformized models to successfully extract unexplored facets of teaching and learning.

The central objective of this paper is to illustrate the potential application of such a framework in the field of humanities. One of the gains of promoting educational designs in light of recent technological developments is that it opens new avenues for the conception of more targeted curriculum designs.

3. Digital Corpora and Learning Analytics

3.1. The argument for digital corpora

Data availability is by far the missing ingredient in the development of responsive educational designs. The most accurate systems use preprocessed and labelled data to maximize the learning of the network. However, this luxury is reserved mostly to languages where curated and labelled digital corpuses exist.

> "Most state-of-the-art machine learning models rely on supervision from large amounts of labeled data—a requirement that cannot be met for the majority of the world's languages (Snyder 2010). Over time, approaches have been developed to address the data bottleneck in multilingual NLP. These include unsupervised models that do not rely on the availability of manually annotated resources (Snyder and Barzilay 2008; Vuli´c,, De Smet, and Moens 2011, inter alia) and techniques that transfer data or models from resource-rich to resource-poor languages (Padó and Lapata 2005; Das and Petrov 2011; Täckström, McDonald, and Uszkoreit 2012, inter alia)."[24]

On the other hand, the scarce resourced languages have the option of the less accurate but more accessible unsupervised neural networks or machine translation. Both the supervised and unsupervised systems require massive amounts of data. If in the recent past, when referring to data it usually meant numbers and quantitative instruments, the current definition adds words and qualitative instruments. This allows full data architects to incorporate hybrid models which target deeper philosophical dimensions of data. While all data communicates behaviors that tell a big story, the qualitative exploration complements the quantitative results with a rich behavioral refinement. The hybrid systems therefore, are increasingly capable to reveal more comprehensive insights on the depth of human nature.

> "Recent AI breakthroughs are based on supervised machine learning. A critical success factor of these systems is the availability of huge amounts of pre-categorized training data. In contrast to logic- and knowledge-based approaches to AI, we therefore characterize these as "data-based" AI systems in this report. Many of these "deep learning" neural AI systems may well be characterized as "datavores." At present, the most important technical bottleneck of AI, therefore, is the availability of data. This is a qualitatively new development in the history of computing and information processing. Without access to vast training datasets, it is very difficult to develop successful AI systems."[25]

Digital corpora is essentially related to and representative of the qualitative dimension. There are a significant number of advantages in implementing digital corpora in research designs. The added benefits include multi perspective text metrics designed to maximize value extraction and insight granularity. As the size of the digital corpora continues to expand, specialised research approaches have been offered a quick refresh. Of the learning analytics sector, corpus based Multidimensional Analysis (MDA) is a quantitative corpus-based approach dedicated to the field of register studies. The MDA approach accounts for linguistic variation that is determined by situational variables, which is to say that, the distribution of linguistic features in a text heavily depends on the register it belongs to (e.g. newspapers, academic lectures, sales pitches, application letters). A technique that identifies groupings of linguistic features by using factor analysis (Biber 1986; ibid. 1988) is extremely helpful in simplifying data related to variation word choice and word length. The dimensions (i.e. clusterings of features) resulting from that analysis are then interpreted according to their communicative function (Conrad 2015: 316-317).[26] Another relevant approach which argues less from a top-down institutional framed classification or clustering standpoint and more from a bottom-up student point of view, is purposed to reflect the learners' perspective. This function is facilitated by both the public and class created corpuses. Data-Driven Learning (DDL) adds authenticity to the class and offers the liberty of guided discovery. Specific to language students, DDL is a pedagogical approach where the direct learner engagement with corpus data, allows users the flexibility to learn and internalise statistical and contextual information about the language in use.

> "DDL allows learners to consult corpus data, query, manipulate and visualise a range of output data including concordances of query words with surrounding context, statistical information in the form of frequency or collocation lists, and increasingly visual or multimodal forms of data. DDL has often been described as an 'inductive' approach to learning, where learners' active engagement with corpus data leads to increased focus-on-form (Long, 1991) and data-enhanced 'noticing' of language features (Schmidt, 1990) that replaces the need to memorise abstract textbook 'rules' while promoting a range of constructivist skills correlated with improved learning practices (e.g. Cobb, 1999)."[27]

Integrating inferential data approaches in platformized architectures shifts the focus from follower to creator of knowledge and personalized learning models.

3.2. The argument for Learning Analytics

MDA and DDL approaches belong to the larger family of EDM and LA, where MDA subscribes to LA and DDL to EDM. Big Data in education has largely emerged in two research areas: Educational Data Mining (EDM) and Learning Analytics (LA). While EDM focuses on developing tools for new pattern discovery in educational data, LA is concerned with the collection, measurement, analysis, visualisation and reporting of data about learners' behavior. In order to optimize both learning and the learning environment, learning analytics can be initially applied to student behavior data retrieved from elearning platforms.

Timeline log data and student digital persona can provide solid grounds for creating comprehensive educational data sets. This approach focuses on the students' online behavior as reflected by the user's digital profile. Observing behavior at scale provides valuable insights about interactions with peers, systems and authority figures. Observational log studies contain partitioning log data by time and by user. Partitioning by time helps understanding the significant temporal features, such as periodicities (including consistent daily, weekly, and yearly patterns) and sharp changes in behavior during important events. It offers an up-to-the-minute map of user interaction by comparing past and current behavior timelines. Partitioning log data by user characteristics shares a pleiade insights as well. Numerous studies grow the research body of literature in the area of log analysis, visualization, clustering and classification. These studies offer a more refined understanding of student academic engagement and disengagement based on online activity inferences. Regarding the discovery and mapping of frequent navigational patterns, other research endeavours turn to learning analytics by employing sequential pattern mining techniques. Log data clustering models are also able to successfully categorize students' learning styles into deep and surface learners. Research interest has also been directed towards detecting the relationship between the student's profile visible variables and student's final grades as well as predicting human activities based on natural language data.[28] The inquiry aims at identifying the impact of a particular activity or set of activities on the final grade. The variety of approaches is continued by Excel macros deployed to analyze and visualize time stamped activities based on the total number of page views, unique users, unique actions, IP addresses, unique pages, average session length and bounce rate. Machine learning architectures focus on the students' topic engagement by evaluating the number of their activities and calculate a grade prediction scenario based on the number and types of engaged activities. From an institutional perspective, by forecasting next years' grade acquisition reflective of students' academic success or dropout rate,[29] instructional designers have the opportunity to make informed decisions on recalibrating the curriculum architecture into a more relevant format. Even if these research strategies were largely applied to Moodle, most elearning platforms can deliver similar value extraction opportunities.

3.3. Measuring the educational act

While paper typographed texts have allowed science to evolve for centuries, the digitalization of writing opens new avenues for academic exploration. More precisely, compared with the traditional ones, the digital text corpuses favor more complex linguistic measurements. With what concerns the teaching and learning as well as the policy design strategies, performance metrics interest seems to be high on the agenda of designing competitive educational architectures on all levels. This is largely facilitated by participative social-constructivist educational concepts operating in flat elearning architectures. In this way, the growth of the digital educational corpora stimulates both the expansion of controlled environments where the experimental modeling takes place and of the diad incorporating the academic writing development as well as the collective student reasoning capabilities. This is generally regarded as a formalized logic design which models on the view that social-constructivist interventions can attain the cognitive goals of general improvement in thinking, reasoning, and problem-solving.[30] The nascence of the educational digital corpus is a modern dimension of linguistics in which the written text becomes both an object of performance measurement but also evidence. While allowing for traceability of both teaching and learning complexities, the user generated corpuses may provide relevant statistical levels of confidence when incorporated in supervised networks. This allows for morphological, syntactic and semantic mapping of the user generated content. The concept banks on the convergence of two different educational research paradigms which account for individual and contextual attributes.

> "One conceptualization is represented by the founders of the emerging discipline of psychology such as Hall (1909) and Thorndike (1906, 1931), who emphasized measuring attitudes, aptitudes, and cognitive abilities of students, then using those measures of individual characteristics to develop tailored interventions to improve the cognitive competencies of individual students. The other major conceptualization comes from the founders of a sociocultural approach to psychology such as Vygotsky (1978) and Cole (Cole, Gay, Glick, & Sharp, 1968), who emphasized the importance of the social context of education and the role of language in developing cognitive competencies.[31] [...] In a seminal paper in 1993, described by Mason in 2007 as "a remarkable event," Pintrich, Marx, and Boyle called for work on conceptual change to include affective, motivational, and situational factors. Attention to such noncognitive variables, and how they might interact with cognitive variables in particular contexts, gave rise to consideration of dual-process models (Dole & Sinatra, 1998). Mason (2007) saw the inclusion of different types of variables as an important step to better understandings of "knowledge restructuring."[32]

The growing availability of digital annotated corpuses signal the increased interest and confidence in the value of text processing extraction potential. As the need to grow the user generated academic digital corpus becomes more evident, an increasing number of text processing instruments become available. The growing list which includes parts-of-speech annotation and tagging, syntax-parsing, lemmatization and synset-annotation are increasingly becoming standardized for English-language corpora.[33] These text processing strategies can be implemented into platformized learning designs to extract valuable insights about the quality of the academic act. Such an approach would reveal important metrics on group and individual interactions, topic sentiment and relevance, gender segmented levels of participation, student satisfaction etc. One of the unique advantages of digitally expanding the academic act, is that it allows for asynchronous users to develop a more personalized and "individual time"[34] of study. Recent deep learning architectures relying on attention detection mechanisms, focus on the lexicons of affective words, examining the relevance of word valence in a given text. It becomes increasingly evident that information about affective words and phrases, and their assigned valence scores provide reliable features for machine learning models. Moreover, when incorporated into deep learning models, such lexical information improves the performance of the model.[35] Depending on the desired depth of the architecture, additional features can be added to include the Big Five personality prediction models. Based on the class forum social interaction data incorporating both multimodal and linguistic features, the network can be set to discover and innfer key personality traits which include openness, conscientiousness, extraversion, agreeableness and neuroticism.[36]

4. The emergence of responsive educational designs

As it was mentioned before, the performance of data mining and depth of learning analytics largely depend on digital corpora. While general digital corpora is useful in training general purpose algorithms, it is the language specific educational corpora that would facilitate best results based on field specific training parameters. As illustrated in the Multi Dimensional Analytics (MDA), higher levels of confidence are associated with results where the data input belongs to the same register. However, for resource poor languages, complying with the educational register and first language conditions remains a monumental task. In terms of general-purpose databases, a tremendous progress has been realised by the Research Institute for Artificial Intelligence "Mihai Drăgănescu" of the Romanian Academy. By successfully compiling the most complete romanian computational corpus to date, the CoRoLA project opens new research avenues by offering public access of structured datasets to research interests targeting linguistic studies, language modeling for automated romanian language processing, development of translation models, semantic classification etc.[37] Attempts to resolve the problem of labelled scarcity in resource poor languages has been tested in a few experimental designs. It has been argumented that machine translation offers a viable alternative to lexicalization in the absence of annotated parallel data. In this model, a source sentence is machine translated into a target language (Banea et al. 2008) through a bilingual lexicon (Durrett, Pauls, and Klein 2012).[38] Other approaches abandon completely the use of annotated resources altogether and focus entirely on unsupervised learning alternatives. Given some latent variables, this class of methods infers probabilistic models based on observations, revealing the hidden structures within unlabeled text data. Although these methods have been used extensively for multilingual applications (Snyder and Barzilay 2008; Vuli´c, De Smet, and Moens 2011; Titov and Klementiev 2012, inter alia), their performance tends to lag behind the more linguistically informed supervised learning approaches (Täckström, McDonald, and Nivre 2013). Encouraging approaches in solving the data scarcity problem include transferring models from resource-rich to resource-poor languages or learning joint models from annotated examples in multiple languages in order to leverage language interdependencies.[39] However transferring linguistic information from rich source languages to resource poor target languages poses some specific challenges.

> "The fact that Romanian verbs are inflected for mood (such as indicative, conditional, subjunctive, presumptive), enables an automatic classifier to identify additional subjective markers in text. Some moods such as conditional and presumptive entail human judgment, and therefore allow for clear subjectivity annotation. Moreover, Romanian is a highly inflected language, accommodating for forms of various words based on number, gender, case, and offering an explicit lexicalization of formality and politeness. All these features may have a cumulative effect in allowing for better classification. At the same time, English entails minimal inflection when compared to other Indo-European languages, as it lacks both gender and adjective agreement [...] Verb moods are composed with the aid of modals, while tenses and expressions are built with the aid of auxiliary verbs. For this reason, a machine learning algorithm may not be able to identify the same amount of information on subjective content in an English versus a Romanian text [...] As shown by some of our experiments, Romanian seems to entail more subjectivity markers compared to English."[40]

It becomes therefore relevant how machine translation loses accuracy when the rich resource language is a minimally inflected language one, as is the case of English. However, the same procedure may yield better results when employing languages belonging to the same language tree such as Italian, Spanish and French. While machine translation may be deployed for general purpose corpus compilation, the reality is that digital educational corpuses are hardly available even in the context of the rich resource languages. Therefore, a long term strategic objective may be formulated in the direction of developing an education based, ideally user generated romanian digital corpus. However, the success of enhanced designs is conditioned by additional remarkably important external dynamics. One chief element is that of challenging the traditional power balance. This comes from the fact that education platforms are known to level both tutors and students as users and contributors, granting not always comfortable levels of advanced measurability on both dimensions. Another challenge relates to the availability of pre trained networks in the language of interest, romanian in this case. While in the end, all of these constructs must demonstrate the economic viability of the approach. This is to say that in terms of data collection, processing, extraction and visualization, such instruments must deliver relevant enough educational process insights in order to offset the investment costs.

Although not always immediately visible, educational platforms subscribe to various philosophies and functional approaches. They can be easily identified in the various turns that science education research has experienced in the recent past. Switching from behaviourist to cognitive (Posner, Strike, Hewson, & Gertzog, 1982) to linguistic (Lemke, 1990) and to the more recent "practice turn" (Ford & Forman, 2006), the shifts encompass both science and technology studies as well as second generation cognitive science.[41] Many if not most platformized educational models operate on prebuilt social markers, which allow for native integration with linguistic models. This process is largely patronized by the fast paced shift in economic and workforce patterns, requiring academic institutions to implement fully responsive programmes that adapt swiftly to changing socio-economic forces. Monitoring teaching and learning parameters can therefore prove very lucrative in maintaining a competitive educational footprint in the educational arena. Fast response times, however, can only be addressed with the help of the state of the art technological advancements. In addition to cloud based applications, the significant increase in CPU and GPU processing power[42] combined with the most recent publicly available datasets,[43] create unprecedented levels of opportunity for educational data mining and innovative curriculum design exploration. This new context enables devices to extract value from traditional educational data silos and visualize them as flow processes. Such a task can be implemented by natively integrating platformized designs with natural language processing models.[44]

5. Conclusions and future directions

McKinsey's Global Institute senior adviser Jacques Bughin, speaking about growing the AI ecosystem in Europe, highlights the fact that "AI is more than technology. As I say, it's about scalability. You need social, emotional skills, you need technical skills, you need digital skills. It's a major transformation, and it's all about the ecosystem."[45] Out of the EU's 28 states, 23 are identified as relevant on the AI readiness index, where Romania holds the last position with average rankings on AI Start-Up, Saving Rate, ICT Connectedness and low rankings on Automation, Digital Readiness, Innovation and Human Skills.[46]

The development of both the general purpose and specialized educational corpora would open new opportunities of data modeling and exploration in the low ranking sectors. For example, one way of stimulating automation, digital readiness and innovation can be initiated by integrating an IPO[47] model into platformized educational designs. In such a framework, the quality of the assigned materials, the student contributions and their relationship with the final grades can be revelatory in terms of educational model performance. By factoring in affect distribution, the design can be considerably enriched on the emotional dimension of teaching and learning processes as well. In exploring the last dimension, seasoned research argues that in computational linguistics, the automatic detection of emotions in texts is becoming increasingly important from an applicative point of view. While emotional granularity is applicable to general interest tasks such as opinion mining, market analysis and affective computing, it is also relevant in the natural language processing interfaces of elearning environments.[48] Particularly to the latter, the affect granularity is set to offer significantly more refined course related sentiment projections. The less observable patterns at play in the educational digital corpora can be further explored with the help of conditional logic architectures. One possible architecture of the IPO model may account for formats where variables such as I = tutor inputs of class assignments, P = tutor - student/s and student - student/s interactions, O = final grades. In a MLP or hybrid MLP-LSTM architecture,[49] the final grades can be cross-referenced against cognitive and affective values from both class interactions and assignments. The cognitive dimension in I and P can be investigated by aggregating academic references ranking in I and keywords in P. Character tagging associated with authors can automatically be parsed in both sections and cross-referenced against topic relevance. Threshold based conditional rules such as, If $c \geq 200$ citations = seminal work or if $c \leq 200$ citations = important work, can be established for citation levels where corpus crawls reveal the differences between important and seminal works. On the affective data exploration task, the architecture can be designed to infer affective patterns. As emotional dynamics usually develop in social interactions, these are very likely to be identified in the P sector. The module can be structured to emulate both Robert Plutchik's eight primary types of emotions[50] and also the negative, neutral, positive, emotional valence classification. The architecture can be designed to evaluate the main boards' discussion threads for emotion and valence, and reposition students based on their similar or different emotional characteristics. While this can be performed on a cognitive basis as well, this feature is particularly useful for pedagogical topic modeling and pedagogical experimental design testing. As the annex A and annex B indicate, text disambiguation elements such as emoji, can greatly influence the valence of the thread, making valence classification more precise. Sentiment calibration can be implemented based on the conditional rule reading that if negative or positive thread valence $\geq 50\%$, then suggest repositioning as new groups based on negative and positive valence features. In terms of cognitive calibration, student threads can be referenced against relevant topic markers. For example if a student mentions Shakespeare notations in the thread, the rule can be optimized to suggest more English literature reading, or, if a student mentions Checkhov then suggest more Russian literature reading assignments. All of these conditional logic features play a critical role in designing time and format sensitive personalized learning environments.[51] The character / author tagging function satisfies both of the register specific and the personalized learning requirement.

While the Romanian text based romanian generated educational corpora must be led by a major initiative to gain traction and stimulate its progress, a faster development and integration is available to non-Romanian text based romanian educational corpora. This scenario includes the romanian academic contributions of native romanian contributors to German, English, Spanish, Portugueze, French and Italian digital corpora.[52] For the task of word alignment through building and using parallel texts,[53] as in the case of Romanian and Spanish, automatic translation represents a viable alternative.[54]

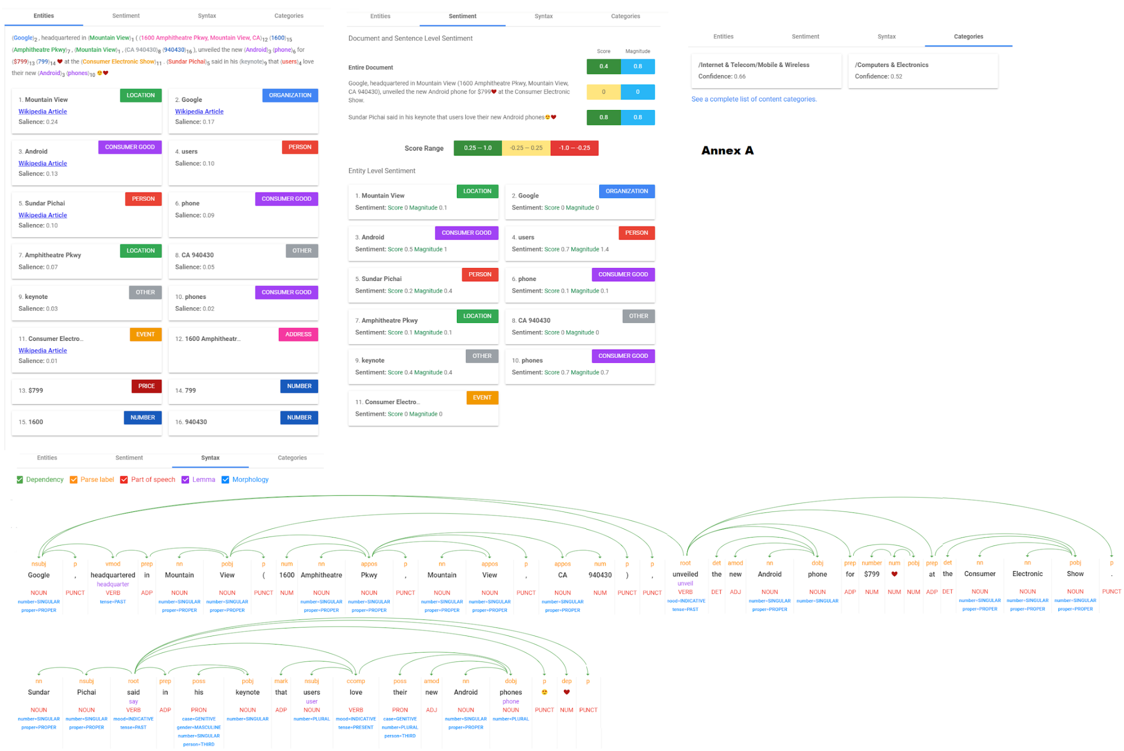
While the potential benefits of integrating education with intelligent architectures is significant, an important number of challenges remain to be addressed. One refers to data exploitation privacy rules while the other points at architecture limitations. AI systems are known to excel at collecting informal evidence of skills, experience, and competence from open and proprietary data sources, including but not limited to social media, learner portfolios and open badges about skills and achievements. This creates both ethical and regulatory challenges.[55] The second area of improvement is related to the fact that "supervised learning" models are based on human labelled training of existing data, which allows them to "see" the world only as a repetition of the past.[56] In supervised AI learning models, the possible choice outcomes need to be provided to the system before it starts to learn. Also, when predictions are made using large samples of historical data, systems' performance may be limited in terms of understanding and accurately processing creativity and innovation. Machine learning designs still need to register significant improvements in the area of event exception identification and representation, which are reflective of breakouts from historical patterns of behaviour. In such a sense, it can be argued that AI has the potential of limiting the expression of human agency.[57] However, the number of benefits visibly outnumber the number of disadvantages. In this context the digitized versions of linguistics are set to gain unprecedented traction and are primed to become strategic assets. As traditional formats reflecting the industrial age link between work and education become increasingly misaligned, new models reflecting the mass adopted technologies are expected to deliver competitive performance. If current educational designs largely address the needs of an outgoing industrial world, the new architectures employ fresh knowledge and data which are created, used, and learned in ways that have not been possible before. Gradually, it becomes evident that AI is not only a solution to problems existing in the current educational systems in the sense of automating repetitive tasks, but also as a solution for future educational architectures.[58]

In the context of platformised learning designs, digital corpora and linguistics signal a rapidly growing level of importance. When integrated with smart device usage it is anticipated that edutech will make written communication more informal, leaving more room for mood and emotional expressivity. The standardized models of human functioning available in mathematical psychology,[59] can further expand the moderately explored potential of both cognitive and affective user generated content. In this way, highly competitive elearning platforms can advance knowledge and offer valuable insights both on the permissive as well as on the prohibitive factors influencing academic achievement.[60]

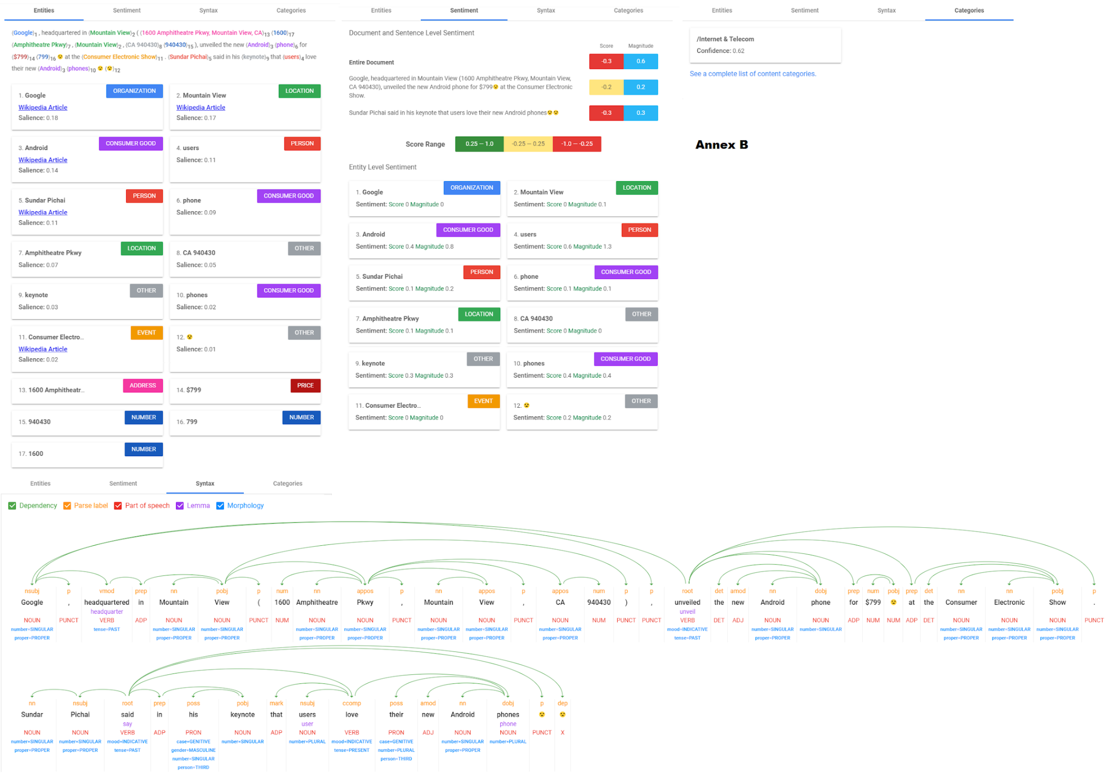
As visualized in Google's NLP API text sample, current text processing models allow for advanced cognitive and emotional measurements. The greatest novelty of such a design rests in its capability of processing real time sentiment analysis jobs, reflective of levels of satisfaction performance. As the case is, along with word level sentiment valence and classification, intelligent designs can account for emoji as emotional disambiguation reinforcers.

As the intelligence of engines is gradually engineered to discreetly operate more complex logics in the back end, the front end is being increasingly tailored for untrained user accessibility. Therefore, the soft emotional and social skills are the upcoming assets of interest expected to handle the finetuning of the intelligent architectures. Whether in its cognitive or affective state, supporting the long-term development of linguistics as digital corpora, may lead to the emergence of a scalable, unified and responsive European educational design.

Annex A



Annex B



Notes:

1. Ilkka Tuomi, "The Impact of Artificial Intelligence on Learning, Teaching, and Education. Policies for the future", in *EUR 29442 EN*, ed. M. Cabrera, R. Vuorikari, and Y. Punie (Luxembourg: Publications Office of the European Union, JRC Science for Policy Report, 2018), 33.
2. Tavis D. Jules, "Big "g" and small "g:" *The Variable Geometries of Educational Governance in an Era of Big Data*", in *The Educational Intelligent Economy: Big Data, Artificial Intelligence, Machine Learning and the Internet of Things in Education*, ed. Florin. D. Salajean, and Tavis Dules (Emerald Publishing Limited: 2020, Vol. 38), 2.
3. Tuomi, "The Impact of Artificial Intelligence," 8.
4. Ibid., 9-10.
5. Ibid., 10.
6. Ibid., 21.
7. Florin D. Salajan, "An Analysis of Rhetorical Devices in Policy Narratives on the European Institute of Innovation and Technology: Implications for European Higher Education," *European Educational Research Journal*, SAGE 2017, 4.
8. https://en.wikipedia.org/wiki/S.M.A.R.T. Self-Monitoring, Analysis and Reporting Technology, accessed on 02.08.2020, at 02.20 am.
9. Tuomi, "The Impact of Artificial Intelligence," 6.
10. https://en.wikipedia.org/wiki/Mathematical_psychology
11. Tuomi, "The Impact of Artificial Intelligence," 28.
12. Ibid., 32.

13. Florin D. Salajan, and Tavis D. Jules, *Introduction: The Educational Intelligent Economy, Educational Intelligence, and Big Data* (North Dakota University, November 2019), 1.
14. https://www.reuters.com/article/us-softbank-group-son/softbanks-son-says-japan-should-make-ai-mandatory-subject-for-college-students-idUSKBN1YL09I, accessed on 01.01.2020, 09.56 am.
15. Salajan, and Jules, *Introduction*, 2.
16. Ibid., 3.
17. Salajan, "An Analysis of Rhetorical Devices," 1-3.
18. Salajan, and Jules, *Introduction*, 1.
19. Mutlu Cukurova, Rosemary Luckin, and Carmel Kent, "Impact of an Artificial Intelligence Research Frame on the Perceived Credibility of Educational Research Evidence," *International Journal of Artificial Intelligence in Education* (December 2019): 1.
20. Tuomi, "The Impact of Artificial Intelligence," 2.
21. Ibid., 5.
22. Cukurova, Luckin, and Kent, "Impact of an Artificial Intelligence Research Frame," 3.
23. Tuomi, "The Impact of Artificial Intelligence," 23.
24. Edoardo Maria Ponti et al., "Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing," *Association for Computational Linguistics*, Vol. 45, Nr. 3 (2019): 560.
25. Tuomi, "The Impact of Artificial Intelligence," 3.
26. Benedikt Szmrecsanyi, and Laura Rosseel, *English Corpus Linguistics: The Current State-of-the-Art, and a Critical Appraisal* (Leuven: Katholieke Universiteit Leuven, 2019), 7.
27. Peter Crosthwaite, *Data-Driven Learning for the Next Generation: Corpora and DDL for Pre-tertiary Learners* (London: Routledge, 2019), 4.
28. Steven R. Wilson, and Rada Mihalcea, *Predicting Human Activities from User-Generated Content* (University of Michigan, arXiv:1907.08540v1 [cs.CL] 19 Jul 2019), 1.
29. Neslihan Ademi, Suzana Loshkovska, and Slobodan Kalajdziski, *Prediction of Student Success Through Analysis of Moodle Logs: Case Study*, in *ICT Innovations 2019, CCIS 1110*, ed. S. Gievska, and G. Madjarov (Berlin: Springer 2019), 28.
30. Paul Webb, and J. W. (Bill) Whitlow, "Merging Cognitive and Sociocultural Approaches: Toward Better Understandings of the Processes of Developing Thinking and Reasoning", in *Theorizing the Future of Science Education Research, Contemporary Trends and Issues in Science Education*, Vol. 49 (Springer Nature Switzerland AG 2019), 18.
31. Ibid., 11.
32. Ibid., 13.
33. Szmrecsanyi, and Rosseel, *English Corpus Linguistics*, 3.
34. Radu Drăgulescu, "Psycholinguistic and Neurolinguistic Approaches on Communicational Distorsions," *The Proceedings of the International Conference Globalization, Intercultural Dialogue and National Identity. Section: Language and Discourse*, (Târgu-Mureș: Arhipelag XXI Press, 2014), 106.
35. Martina Toshevska, and Slobodan Kalajdziski, "Exploring the Attention Mechanism in Deep Models: A Case Study on Sentiment Analysis," in *ICT Innovations 2019, CCIS 1110*, ed. S. Gievska, and G. Madjarov (Berlin: Springer, 2019), 202, 210.
36. Filip Despotovski, and Sonja Gievska, "An In-Depth Analysis of Personality Prediction", in *ICT Innovations 2019, CCIS 1110*, ed. S. Gievska and G. Madjarov (Berlin: Springer, 2019), 134-135. (FFM-OCEAN).
37. http://www.racai.ro/p/corola/, accessed on 14.02.2020 at 01.27 am.
38. Ponti et al., "Modeling Language Variation," 566.
39. Ibid., 564.
40. Carmen Banea, Rada Mihalcea, Janyce Wiebe, and Samer Hassan, "Multilingual Subjectivity Analysis Using Machine Translation", *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (Honolulu: Association For Computational Linguistics, October 2008): 134.
41. Kok-Sing Tang, "Scientific Practices as an Actor-Network of Literacy Events: Forging a Convergence Between Disciplinary Literacy and Scientific Practices," in *Theorizing the Future of Science Education Research, Contemporary Trends and Issues in Science Education* Vol. 49, ed. V. Prain, and B. Hand (Springer Nature Switzerland AG 2019), 83-84.
42. https://www.amd.com/en/products/cpu/amd-ryzen-threadripper-3990x#product-specs, accessed on 07.02.2020 at 10.00 pm.
43. https://towardsdatascience.com/google-just-published-25-million-free-datasets-d83940e24284, accessed on 08.02.2020 at 09.45 am.
44. https://cloud.google.com/natural-language/, accessed on 08.02.2020 at 09.30 pm.
45. Khary Johnson, "How Europe's AI ecosystem could catch up with China and the U.S." October 1 2019 8:32 AM https://venturebeat.com/2019/10/01/how-europes-ai-ecosystem-could-catch-up-with-china-and-the-u-s/, accessed on 01.01.2020, 10.15 am.
46. Jacques Bughin, "Tackling Europe's gap in digital and AI," February 2019, Discussion Paper, https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-europes-gap-in-digital-and-ai, accessed on 01.01.2020, 10.26 am.

47. Input, Process, Output.
48. Carlo Strapparava, and Rada Mihalcea, "Learning to Identify Emotions in Text," SAC'08 March 1620 (Fortaleza, Cear´a, Brazil, 2008), 1, 2-3; Carlo Strapparava, and Rada Mihalcea, "SemEval-2007 Task 14: Affective Text," *Proceedings of the 4th International Workshop on Semantic Evaluations* (Prague: Association for Computational Linguistics, June 2007), 70; Alexandra Balahur, Rada Mihalcea, and Andrés Montoyo, "Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications," *Computer Speech and Language 28* (2014): 1.
49. Multi Layer Perceptron - Long Short Term Memory design.
50. Mohsin Manshad Abbasi, and Anatoly Beltiukov, "Summarizing Emotions from Text Using Plutchik's Wheel of Emotions, Advances in Intelligent Systems Research," Vol. 166, 7th *Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS)* (Paris: Atlantis Press, 2019), 291 (e.g. anger, fear, sadness, disgust, surprise, anticipation, trust, and joy).
51. Radu Drăgulescu, *Trends and new technologies in teaching and learning General Linguistics. Edmodo, in Intercultural Exchanges in the age of globalization* (Saarbrücken: LAP Lambert Academic Publishing, 2015), 574.
52. Radu Drăgulescu, "Online Media and New Technologies in Teaching Linguistic Disciplines," *Proceedings of the International Conference Globalization, Intercultural Dialogue and National Identity* (Târgu-Mureș: Arhipelag XXI Press, 2014), 140.
53. Joel Martin, Rada Mihalcea, and Ted Pedersen, "Word Alignment for Languages with Scarce Resources," *Proceedings of the ACL Workshop on Building and Using Parallel Texts, Association for Computational Linguistics* (Ann Arbor, June 2005), 6.
54. Carmen Banea, Rada Mihalcea, Janyce Wiebe, and Samer Hassan, "Multilingual Subjectivity Analysis Using Machine Translation," *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (Honolulu: Association For Computational Linguistics, Honolulu, October 2008), 127.
55. Tuomi, "The Impact of Artificial Intelligence," 4.
56. Ibid., 15-16.
57. Ibid., 36.
58. Ibid., 34-35.
59. Eric Schulz, Maarten Speekenbrink, and Andreas Krause, "A Tutorial on Gaussian Process Regression: Modelling, Exploring, and Exploiting Functions," *Journal of Mathematical Psychology* no. 85 (2018), 1.
60. Radu Drăgulescu, "Qualitative Research on Learning Romanian as a Foreign Language in Endo-Linguistic Context," *Revista Transilvania* (Ianuarie 2019): 75.

Bibliography:

Abbasi Manshad Mohsin, and Anatoly Beltiukov. "Summarizing Emotions from Text Using Plutchik's Wheel of Emotions." *Advances in Intelligent Systems Research*, Vol. 166, 7th Scientific Conference on Information Technologies for Intelligent Decision Making Support (ITIDS), 291 - 294. Paris: Atlantis Press, 2019.

Ademi, Neslihan, Suzana Loshkovska, and Slobodan Kalajdziski. "Prediction of Student Success Through Analysis of Moodle Logs: Case Study." In *Communications in Computer and Information Science 1110, Big Data Processing and Mining, 11th International Conference, Ohrid, North Macedonia, October 17–19*, edited by S. Gievska and G. Madjarov, 27-40. Berlin: Springer, 2019.

Crosthwaite, Peter. *Data-driven learning and younger learners: Introduction to the volume. In Crosthwaite (Ed.) Data-Driven Learning for the Next Generation: Corpora and DDL for pre-tertiary Learners*. London: Routledge, 2019.

Despotovski, Filip, and Sonja Gievska. "An In-Depth Analysis of Personality Prediction." *Communications in Computer and Information Science 1110, Big Data Processing and Mining, 11th International Conference, Ohrid, North Macedonia, October 17–19*, edited by S. Gievska and G. Madjarov, 134-147. Berlin: Springer 2019.

Balahur, Alexandra, Rada Mihalcea, and Andrés Montoyo. *"Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications." Computer Speech and Language* 28 (2014): 1-6.

Banea, Carmen, Rada Mihalcea, Janyce Wiebe, and Samer Hassan. "Multilingual Subjectivity Analysis Using Machine Translation." *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, 127-135. Honolulu: Association For Computational Linguistics, October 2008.

Cukurova, Mutlu, Rosemary Luckin, and Carmel Kent. "Impact of an Artificial Intelligence Research Frame on the Perceived Credibility of Educational Research Evidence." *International Journal of Artificial Intelligence in Education* no. 30 (December 2019): 205-235.

Drăgulescu, Radu. "Online Media and New Technologies in Teaching Linguistic Disciplines." *Proceedings of the International Conference Globalization, Intercultural Dialogue and National Identity*. Târgu-Mureș: Arhipelag XXI Press, 2014.

Drăgulescu, Radu. "Psycholinguistic and Neurolinguistic Approaches on Communicational Distorsions." *Proceedings of the International Conference Globalization, Intercultural Dialogue and National Identity*. Târgu-Mureș: Arhipelag XXI Press, 2014.

Drăgulescu, Radu. "Qualitative Research on Learning Romanian as a Foreign Language in Endo-Linguistic Context." *Revista Transilvania* no. 1 (2019): 73-81.

Drăgulescu, Radu. “Trends and new technologies in teaching and learning General Linguistics. Edmodo.” *Intercultural Exchanges in the age of globalization,* 568-575. Saarbrücken: LAP Lambert Academic Publishing, 2015.
Drăgulescu, Radu. “Observații Privind Anxietatea Învățării Limbii Române ca Limbă Străină și Comunicarea Interculturală.” *Revista Transilvania* no. 2 (2019): 84-90.
Drăgulescu Radu. “Considerații privind Statutul Limbii Române ca Limbă Maternă, Limbă Secundară și Limbă Străină.” *Revista Transilvania*, no. 11-12 (2017): 83-89.
Graham, Steve. “Writers in Community Model: 15 Recommendations for Future Research in Using Writing to Promote Science Learning.” In *Theorizing the Future of Science Education Research*, edited by V. Prain, and B. Hand. Switzerland: Springer Nature AG, 2019.
Ilkka, Tuomi. *The Impact of Artificial Intelligence on Learning, Teaching, and Education. JRC Science for Policy Report.* Edited by Marcelino Cabrera, Riina Vuorikari, and Yves Punie, 1-42. Luxembourg: Publications Office of the European Union, JRC Science for Policy Report, 2018.
Joel, Martin, Rada Mihalcea, and Ted Pedersen. “Word Alignment for Languages with Scarce Resources.” *Proceedings of the ACL Workshop on Building and Using Parallel Texts*. Ann Arbor, MI, Association for Computational Linguistics (June 2005): 65–74.
Ponti, Edoardo Maria et al. “Modeling Language Variation and Universals: A Survey on Typological Linguistics for Natural Language Processing”. *Computational Linguistics* 45 (6) (2019): 1-43.
Salajan, Florin D., and Jules D. Tavis. *Introduction: The Educational Intelligent Economy.* In *Educational Intelligence and Big Data in The Educational Intelligent Economy: Big Data, Artificial Intelligence, Machine Learning and the Internet of Things in Education*, edited by Florin D. Salajan, and Jules D. Tavis. Bingley: Emerald Publishing Limited, 2020.
Schulz, Eric, Maarten Speekenbrink, and Andreas Krause. “A tutorial on Gaussian Process Regression: Modelling, Exploring, and Exploiting Functions.” *Journal of Mathematical Psychology,* no. 85 (2018): 1-16.
Strapparava, Carlo, and Rada Mihalcea. “SemEval-2007 Task 14: Affective Text.” *Proceedings of the 4th International Workshop on Semantic Evaluations*. Prague: Association for Computational Linguistics, 2007.
Szmrecsanyi, Benedikt, and Laura Rosseel. *English Corpus Linguistics: The Current State-of-the-Art, and a Critical Appraisal*. Leuven: Katholieke Universiteit Leuven, 2019.
Tang Kok-Sing. “Scientific Practices as an Actor-Network of Literacy Events: Forging a convergence Between Disciplinary Literacy and Scientific Practices.” In *Theorizing the Future of Science Education Research, Contemporary Trends and Issues in Science Education Vol. 49*, edited by V. Prain, and B. Hand 83-98. Switzerland: Springer Nature, 2019.
Toshevska, Martina and Slobodan Kalajdziski. *Exploring the Attention Mechanism in Deep Models: A Case Study on Sentiment Analysis.* In *Communications in Computer and Information Science 1110, Big Data Processing and Mining, 11th International Conference, Ohrid, North Macedonia, October 17–19*, edited by S. Gievska and G. Madjarov. Berlin: Springer 2019.
Webb, Paul, and Bill J. W. Whitlow. “Merging Cognitive and Sociocultural Approaches: Toward Better Understandings of the Processes of Developing Thinking and Reasoning.” In *Theorizing the Future of Science Education Research, Contemporary Trends and Issues in Science Education*, edited by V. Prain, and B. Hand. Switzerland: Springer Nature, 2019.
Wilson, Steven R., and Rada Mihalcea. “Predicting Human Activities from User-Generated Content.” *arXiv:1907.08540v1 [cs.CL]* (19 Jul 2019): 1-11.