Analyzing the Discourse of Chais Conferences for the Study of Innovation and Learning Technologies via a Data-Driven Approach

Vered Silber-Varod, Yoram Eshet-Alkalai, and Nitza Geri
The Open University of Israel, Raanana, Israel

vereds@openu.ac.il; yorames@openu.ac.il; nitzage@openu.ac.il

Abstract

The current rapid technological changes confront researchers of learning technologies with the challenge of evaluating them, predicting trends, and improving their adoption and diffusion. This study utilizes a data-driven discourse analysis approach, namely culturomics, to investigate changes over time in the research of learning technologies. The patterns and changes were examined on a corpus of articles published over the past decade (2006-2014) in the proceedings of Chais Conference for the Study of Innovation and Learning Technologies – the leading research conference on learning technologies in Israel. The interesting findings of the exhaustive process of analyzing all the words in the corpus were that the most commonly used terms (e.g., pupil, teacher, student) and the most commonly used phrases (e.g., face-to-face) in the field of learning technologies reflect a pedagogical rather than a technological aspect of learning technologies. The study also demonstrates two cases of change over time in prominent themes, such as "Facebook" and "the National Information and Communication Technology (ICT) program". Methodologically, this research demonstrates the effectiveness of a data-driven approach for identifying discourse trends over time.

Keywords: mapping topics in the domain of learning technologies, academic discourse of learning technologies, culturomics, quantitative text analysis, data-driven discourse analysis, word frequency analysis

Introduction

In a world characterized by rapid technological changes, learning technologies researchers face a paradoxical situation in which they are examining how teachers and students are coping with frequently changing learning environments (Cuban, 1999). The paradox lies in the increasing rate of change in learning technologies within the education system, which makes optimal implementation difficult as well as conducting in-depth research on the effectiveness of utilizing these technologies.

Cuban (1999) refers to this technological paradox in the educational arena, claiming that experts have urged teachers to teach Information and Communication Technology (ICT) skills, such as using the Internet, e-mail, producing multimedia projects, and ap-

plications relevant to the constantly changing workplace. These demanding requests of experts led to the paradox of "increasing access and infrequent use" of technologies by teachers (Cuban, 1999, p. 68). In another article, Cuban, Kirkpatrick, and Peck (2001) explain that this paradox led to another one, namely that "new technologies will, paradoxically, sustain old practices" (Cuban et al., 2001, p. 830). Similarly, Eshet-Alkalai and Chajut (2009) showed that educators are dealing with a mixed effect of digital literacy over time: on the one hand, experience and training improve learners' performance with digital technologies, but on the other hand, for the skills that require critical and creative thinking (i.e., creative use of text and critical thinking about information), experience and exposure to information seem to have a negative effect on the users' performance (Eshet-Alkalai & Chajut, 2009). Research findings such as these reflect the difficulties in managing systemic decision-making on learning technologies, let alone research, in a world of fast-changing technologies.

The aim of the current study is to examine the nature of changes over time in learning technologies. We assume that a detailed longitudinal text-analysis of proceeding publications will reveal patterns of change through time in the major areas of research, theories, concepts, pedagogic approaches and research methodologies of learning technologies.

Background

Retrospective investigations of change trends in learning technologies are usually based on governmental reports, such as in Culp, Honey, and Mandinach (2005) or Bakia, Murphy, Anderson, and Trinidad (2011), the latter described the profile of ICT in the Israeli education system (Bakia et al., 2011, pp. 195-201). In a study on ICT in Israel, Elgali and Kalman (2011) present a qualitative analysis of twelve reports prepared by national educational ICT policy-making committees over three decades (1980-2010), as well as of interviews with key participants of these committees. Their longitudinal analysis illustrated the tendency of those committees to recommend adoption of "fashionable" and short-lived technologies, which led to frequent changes within relatively short time periods. Consequently, some technologies or pedagogies exhibited a "spiral" behavior, in which they appeared and disappeared over short periods of time (Elgali & Kalman, 2011).

Recently, scientometrics, which is the study of science, technology, and innovation from a quantitative perspective (Leydesdorff & Milojevic, 2015), was found to be useful for measuring the relationships between trends of change in learning technologies and research of learning technologies (Raban & Gordon, 2015; Van den Besselaar & Heimeriks, 2006). Using data-driven discourse analysis, we examine the patterns and changes as they are reflected in the texts of the articles published over the past decade (2006-2014) as part of the Chais Conferences for the Study of Innovation and Learning Technologies. Chais Conference is the main conference on this topic in Israel and is organized by the Center for the Study of Innovation in Learning Technologies at the Open University of Israel. Chais Conference is the place where leading researchers and developers in the field of innovation and learning technologies in Israel meet and present their research. The main themes of the conference in the past decade have been (Geri, Blau, Caspi, Kalman, Silber-Varod, & Eshet-Alkalai, 2015):

- Empirical examination of models of learning in technological environments;
- Study of the cognitive aspects of learning in technological environments;
- Examination of the effectiveness of teaching using content-driven applications;
- Study of the interactions between learner and computer;
- Technology as a bridge to overcoming distance and gaps;
- Study of the integration of technology into school and higher education systems in Israel;
- Examination of innovative paradigms in instruction and learning in organizations and industry;

- Learning technologies to advance people with special needs;
- Groundbreaking learning technologies;
- Innovative models of online learning;
- Open education resources; massive open online courses (MOOCs);
- Lifelong learning; learning technologies and the reduction of social gaps;
- Digital games as learning environments; crowdsourcing;
- Interaction on social networks;
- Economic, organizational, legal and ethical aspects of integrating learning technologies;
- Measuring the added value of new technologies and of information management in organizations and institutions;
- Resistance to adoption of new technologies in organizations and institutions.

Some of the themes have accompanied the research discourse throughout the nine years covered by this study, while others first appeared somewhere along the way. The conference proceedings reflect the state-of-the-art research in this field in Israel (Geri et al., 2015). Hence, discourse analysis of the Chais Conference articles may shed light on trends in the development of research in the field of learning technologies in Israel and provide answers to the following questions: What are the trends of change in the foci of learning technologies research throughout the years? What are the trends of change within the prominent themes? Which words represent the fundamental terms of this research field?

The current study demonstrates a novel use of a data-driven approach to analyzing trends in academic and research discourse regarding learning technologies. We aim to trace those features in the field of learning technologies as they are expressed in the several topics mentioned above, and to examine whether it is possible to identify trends of change within this area of research.

A Data-Driven Approach to Discourse Analysis

In the data-driven approach, researchers base their work on natural empirical data that derive directly from the studied corpus. In the past decades, a linguistic field known as 'corpus linguistics' has emerged (Sinclair, 1991, 2004). In corpus linguistics, in addition to the identification of structures within the linguistic system, the use of language is mapped in the sense of "trust the text", as Sinclair (2004) entitled his book – *Trust the text: Corpus and discourse*. Researchers are using a corpus-based analysis to describe the features of a particular language or genre. Johnstone (2008) applied such a heuristic approach to discourse analysis. She claimed that discourse analysis is a "systematic, rigorous way of suggesting answers to research questions posed in and across disciplines" (Johnstone, 2008, p. xiii).

In order to illustrate the analytical power of the corpus-driven approach, Michel et al. (2011) created a data set based on 5,195,769 digitized books dating between 1800 and 2000, accounting for approximately 4% of all books ever published in English (Michel et al., 2011, p. 176). Their analysis used a data-mining method, known as the n-gram, to measure cultural trends as expressed in linguistic choices in the texts themselves. An interesting example of research use of the data-driven approach of discourse analysis can be found in Johnstone (2008), whose study examined naming and terminology in relevant articles and revealed how researchers define long-term processes such as aging versus the necessity of special needs (Johnstone, 2008, p. 7).

The accelerated use of the corpus linguistics approach was the result of the development of sophisticated algorithms for natural language processing. The use of advanced statistical methods to analyze the data of large-scale textual databases makes it possible to identify patterns of change and to encode essential qualities expressed in the texts.

One of the spin-offs of corpus linguistics is called 'culturomics' (Michel et al., 2011). Culturomics quantitatively investigates massive digital arrays of written text and spoken language in order to examine cultural patterns in various disciplines (Bohannon, 2011). Culturomics also makes it possible to monitor trends of change and identify the unique lexical items in a particular field by analyzing the publications in that field (Soper & Turel, 2012a, 2012b; Soper, Turel, & Geri, 2014).

Another application of quantitative analysis of texts which has been gaining momentum in recent years is the semi-automatic assessment of exams (Bennet & Ben-Simon, 2005). This application was developed at Israel's National Institute for Testing and Evaluation (Ben Simon & Safran, 2012). Following the development of corpus linguistics, automatic and semi-automatic tools that facilitate and accelerate analysis have appeared, such as Matrix (Rayson, 2003) and AntConc (Anthony, 2011). The use of such tools enables observation of an entire body of knowledge (the corpus) and not just certain elements of its features (Van den Besselaar & Heimeriks, 2006). In the era of Big Data, scholars from different disciplines have adopted diverse computational methods that use word frequency levels and the ratios between different levels of word frequency to make educated forecasts in economics and politics – for example on online news websites (Radinsky, Agichtein, Gabrilovich, & Markovitch, 2011; Radinsky & Horvitz, 2013).
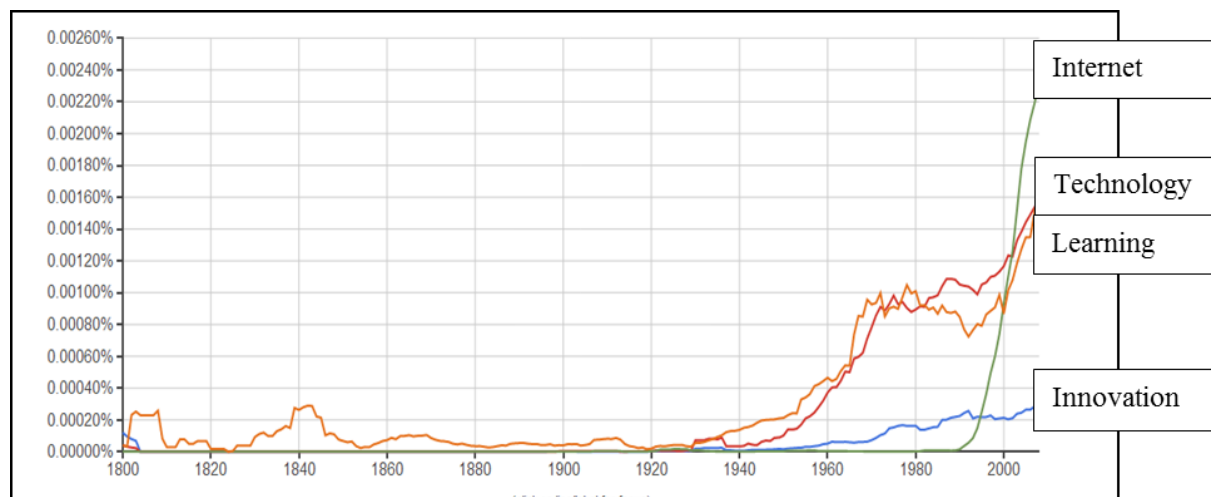

Figure 1. Frequency of use of four terms in the field of learning technologies – internet, technology, learning, and innovation – across two centuries, as retrieved from the database of books in Hebrew, using the Google Books Ngram Viewer platform

One of the requirements for a data-driven discourse analysis is that the chosen corpus must be representative of the topic under study (Livnat, 2012). In order to examine the potential of this kind of analysis for identifying trends of change over time in the field of learning technologies, we conducted a test of the Google Books Ngram Viewer (http://books.google.com/ngrams/info) in Hebrew. This is an open, diachronic database of books. Figure 1 shows the trends of change in the appearance of the words "learning", "internet", "technology" and "innovation" in Hebrew literature over the past 200 years (1800-2008) and the frequency ratios between them. The figure demonstrates the inherent potential of textual analysis, not only in identifying patterns of change over time for individual words, but also in identifying the appearance and disappearance of a word. Thus, for example, Figure 1 shows that the word "learning" was already in use in the 19th century (and presumably even earlier). The word "technology" also appeared in the 19th century, but its use increased significantly as of the 1930s, parallel to the appearance of the early television at the end of the 1920s. The word "innovation" starts to appear in the 1950s, whereas "inter-

net" literally breaks through in the 1990s. Likewise, Figure 1 indicates the simultaneous growth of use of both "learning" and "technology".

Another study that applied a data-driven approach to identify trends in learning technologies research through the analysis of articles published on Google Scholar between 2000 and 2014, found that 3-7 years pass between the appearance of a technology and its comprehensive discussion in the research literature (Silber-Varod, Eshet-Alkalai, & Geri, 2016).

Research Method

The question examined in this study is: what are the trends of change over time in the foci of learning technologies research as reflected in the keywords and full text of the corpus of articles published at the Chais Conferences for the Study of Innovation and Learning Technologies research between 2006 and 2014? The study adopted a data-driven approach as described above. The encoding of the content was informative: to each word we attached the year of its appearance and the article it appeared in.

Corpus Data

The dataset for the study consisted of 1,407 keywords which were assigned by the authors of the 553 articles and poster abstracts (henceforth, both complete articles and poster abstracts are referred to as articles) of the Chais Conferences held between 2006 and 2014 (383 in Hebrew and 170 in English), as well as the full text of these articles.

Analysis of the full text of the articles was conducted through two methods. First, the generation of lists of words, as will be explained below, and second, examination of sequences of words in the articles. Examination of the frequency of sequences was conducted using the n-gram method. An n-gram refers to a sequence of words of length N. An n-gram analysis is based on calculating the relative frequency that a certain n-gram appears in a dataset (Soper & Turel, 2012a). This method examines the frequency of each word sequence, not only common phrases, or well-known collocations such as "high school" or "higher education", but also of any two or more words that appear in sequence in a text (such as the sequence "a significant difference was found", which appears in the dataset of the Chais Conferences 65 times in 39 different articles).

Word List Generation Process

One of the most common methods for processing information from texts is to prepare frequency lists for single words. However, this method of natural text analysis has many limitations. For example, in Hebrew, relying on the counting of only written words, i.e., any sequence of letters (grapheme) between spaces, will be misleading since *ha-lemida* "the learning" and *ba-lemida* "for learning" (each a single grapheme, since articles and prepositions are affixed to the noun in Hebrew) for example, will be counted as two different words. Hence, the unit of analysis in which natural language is processed, i.e., a text, must be the *lemma* – the basic form of the word. This means a process of uniting words that may differ in declension or conjugation but have the same meaning and are from the same lexical category. For example, in Hebrew the lemma *talmid* (student, [masc.]) is also the basic form for the feminine form (*talmida*), the plural form (*talmidim*), the possessives (e.g., *talmidav* – his students) and so forth. In other words, lemmatization unites words with prefixes and suffixes that perform various lexical and grammatical functions in Hebrew.

In this study, we performed the conversion of the words from their written form into their lemmas with the help of the National Institute for Testing and Evaluation, which has advanced tools for morphological analysis of Hebrew (for an overview of the tools comprising the processing system, see Ben-Simon & Cohen, 2011; Ben-Simon & Safran 2012).

Preparing the lists of words involved removing titles and sub-titles, authors' names, reference lists, figures, and tables, leaving only the body of the text to undergo the lemmatization explained above.

Initially, textual analysis was conducted separately for articles in Hebrew and in English, and later the findings from both analyses were manually unified in the process described in Table 1.

Table 1. Database of Chais articles 2006-2014

	Hebrew	**English**	**Full dataset**
Stage 1: Dataset based on Chais Conference articles published in the Proceedings for the years 2006-2014 (See Appendix A for list of conference proceedings and references)			
	480,306 tokens* 42,008 types**	249,769 tokens 10,886 types	730,075 tokens
Stage 2: Creating lists of words from the articles. The Hebrew dataset underwent lemmatization and function words were removed from the English dataset			
	382,941 tokens 12,310 types	124,237 tokens 10,431 types	507,178 tokens
Stage 3: Reduction of the dataset to: nouns, adjectives and proper nouns (e.g., learning, technology, education, collaborative, Facebook), and reduction of the list to 549 words in Hebrew with a frequency of 100 occurrences and above over the years. Integration of English words into the Hebrew list was conducted manually.			
	226,619 tokens 549 types	60,499 tokens 506 types	287,118 tokens

* Tokens = total number of occurrences of a single word in the corpus
** Types = the number of different words appearing in the corpus

Figure 2 illustrates the process of reducing the list of words as a result of lemmatization and of elimination of words with very low frequency. The recall ratio from the start of the process to its completion is 39%.
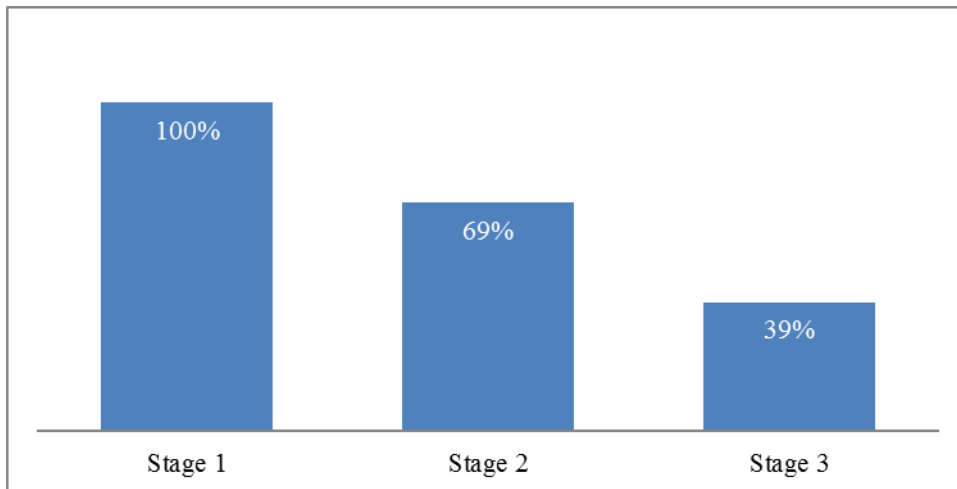

Figure 2: Illustration of the reduction process of the final word list that was used for the analysis

Data Processing and Analysis

We conducted the following tests on the corpus created in Stage 3 above: identification of the fundamental terms in the field of learning technologies; identification of the prominent words in each year; calculation of the dimension of change over time.

First of all, we determined the words which had a global frequency of over 2000 appearances over the years. With this method we expected to identify the words that appear a large number of times each year. A fundamental word was defined as one that was found at a distance of no more than a single standard deviation from its average relative frequency.

Prominent words are unusually frequent (or infrequent) words in the corpus in comparison with the words in a reference corpus. This allows us to identify characteristic words in the studied corpus. In the present study, calculation of the prominence of a word was carried out as a comparison of the frequency in a particular year as opposed to all the preceding years, using a statistical calculation known as log-likelihood (Rayson & Garside, 2000), which enables comparison between relative frequencies of words from different corpora.

Calculation of the dimension of change over time was done by comparing the log-likelihood data for each of the years 2011-2014 to the preceding years.

Findings

This section presents the findings, while distinguishing between trends that emerged from the analysis of the keywords to those that emerged from analysis of the full texts.

In the first part we present an analysis of the frequencies conducted on the keywords given by the authors of the articles.

In the second part we present first the findings for sequences at the level of 3gram found in the articles. We then present the keywords of the field of learning technologies and innovation as found in the word list generation process described above. We present results for frequency, relative frequency, average and standard deviation. Finally, we present the prominent words as of the year 2011 which were found through the log-likelihood calculations compared with the list of words for the year examined as opposed to the list of words for all the preceding years.

Keywords

The dataset consisted of 1,407 keywords assigned by the authors of the 553 articles. The most frequent keyword in the examined corpus was "collaborative learning" (40 instances). This was followed by "distance learning" and "E-learning" in 25 and 23 of the articles respectively. "Higher education", "Wiki", "ICT", "social network", "teacher training", "Facebook" and "learning" were next with a frequency of appearance in only between 18 and 11 articles (Figure 3). The remaining 1,398 keywords appeared in 10 articles or less. Most of the items (1,061 keywords) were singletons. Accordingly, the frequency distribution presented in Figure 3 for all the 1,407 keywords (presented as key term codes in Figure 4) for the years 2006-2014, have a very "long tail" (Zipf's law), which is a distribution typical of a natural textual corpus (Rayson, 2003, p. 36).
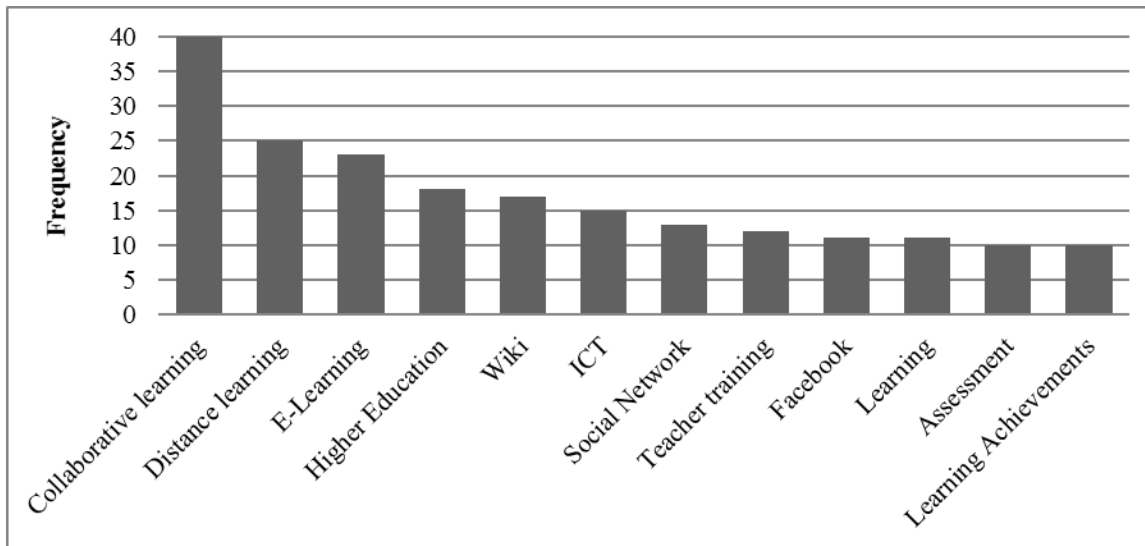
Figure 3. The most frequent keywords (>= 10 occurrences) in the proceedings, 2006-2014
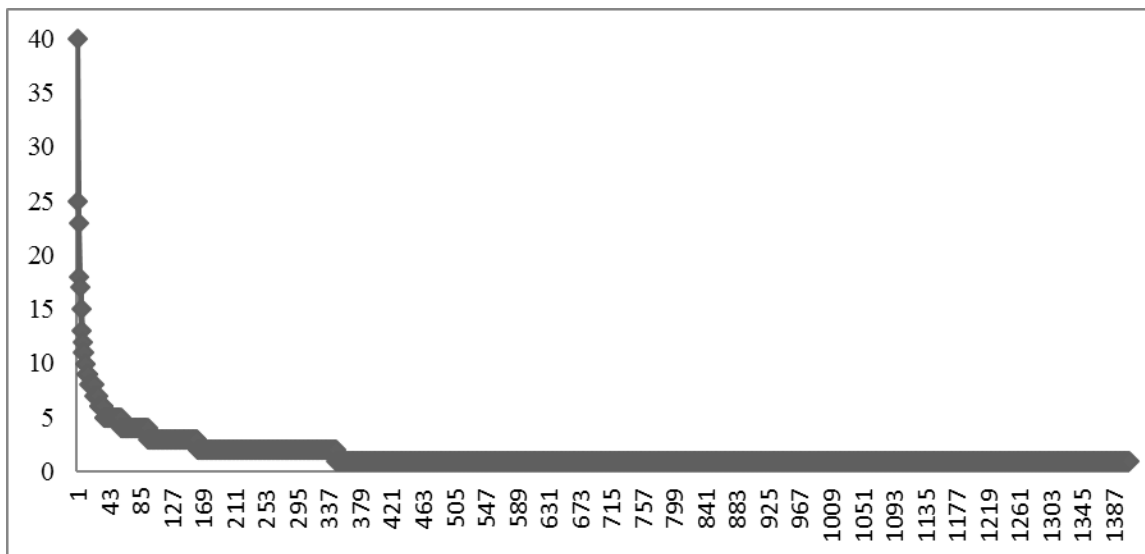

Figure 4. Frequency distribution of the 1,047 keywords, 2006-2014

Word Sequences

This test examined 614,035 different word sequences, and here too, their frequency distribution followed Zipf's law (less than 0.5% of the sequences appear 10 or more times). To the frequency data we added that dispersion data across the articles. For example, when examining sequences that appeared over 50 times and for a spread of more than 50 articles (about 9% of all the articles), we found only five different sequences. The main findings that integrate the data for frequency and spread are the most common sequence in the dataset was "face-to-face" (356 instances in 113 articles). The next 20 most frequent sequences are almost all typical of scientific writing (e.g., "found significant differences") rather than relating to content. However, the sequence in the 14th place is "teachers at schools" (95 instances in 11 articles) and in the 20th place – "higher order thinking", which appears 39 times in 17 articles.

Fundamental Terms in Chais Articles – The Building Blocks of Learning Technologies Research

Contrary to the two previous sections dealing with keywords and natural sequences in the text, in this section we present findings relating to the lists of words generated in the lemmatization process.

Table 2 shows the 18 most frequent words found in the articles, as well as the average and standard deviation of the relative frequencies for each year over nine years. As explained above, we determined the fundamental terms, which had a global frequency of over 2000 appearances over the years. A *fundamental word* was defined as one that was found at a distance of no more than a single standard deviation from its average relative frequency.

Our findings show that the word "research" is the most common, and is exceptional in that it reflects the fact that this is a corpus of research articles, as opposed to the other words, which reflect the objects of the research. A look at these words shows that they belong to the domain of pedagogy and not to that of learning technologies.

Table 2. Relative frequency averages and standard deviations for the 18 fundamental words in research on learning technologies as reflected in the Chais articles, 2006-2014

Lexeme	Average (%)	S.D. (%)
research	2.041	0.177
learning	1.888	0.240
student	1.902	0.442
teacher	1.471	0.503
pupil	1.381	0.436
teaching	1.103	0.199
usage	1.069	0.142
group	1.045	0.325
process	0.925	0.240
information	0.848	0.192
course	0.861	0.279
environment	0.816	0.183
home	0.769	0.306
book	0.748	0.293
technology	0.753	0.116
knowledge	0.757	0.140
system	0.752	0.269
activity	0.722	0.174

Figure 5 shows the relative frequency for each year of the words in Table 2. The data on the Y axis is presented in a log scale, since the log values are normalized and enable presentation of all 18 words within a range of log 1 (between -0.5 and 1.5). The figure illustrates the stability (lack of fluctuation) of the appearance of these words across the nine years studied.
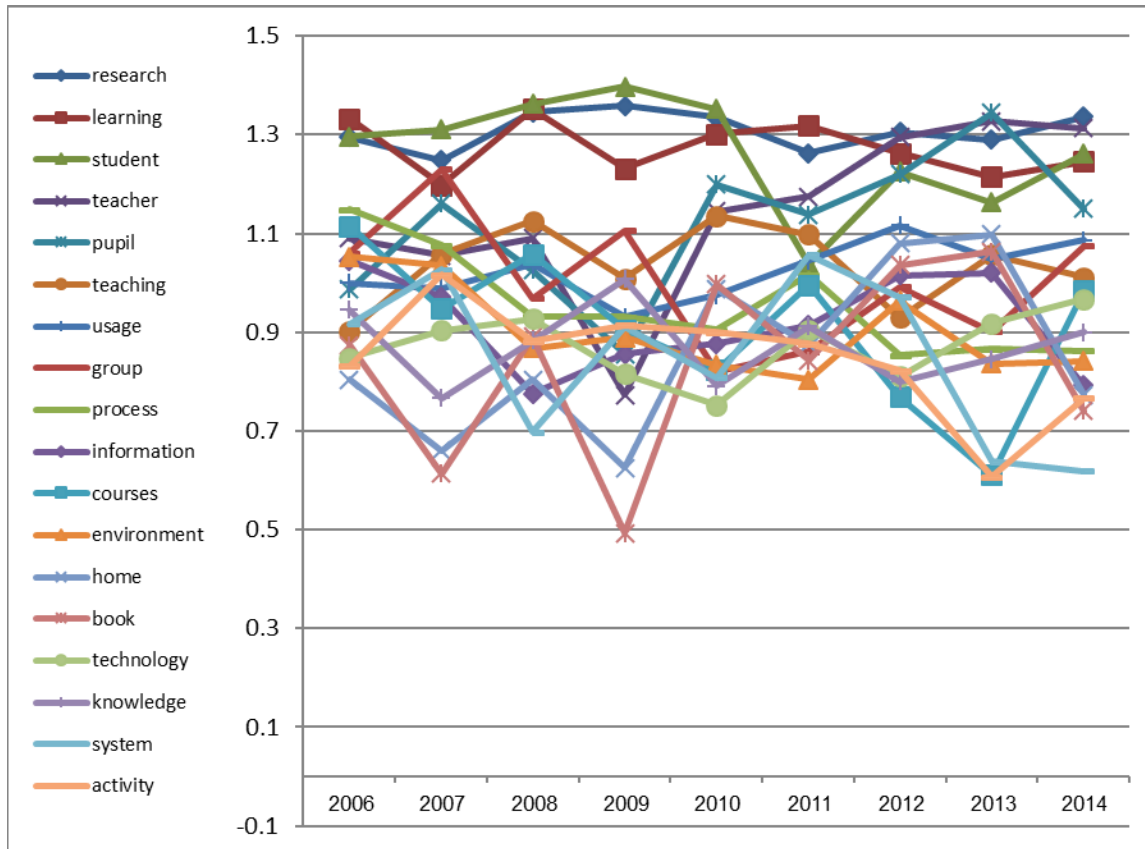

Figure 5. Relative frequency (on a log scale between 0.5-1.5) of the 18 fundamental words in research in learning technologies as reflected in the Chais articles (2006-2014)

Particularly High Frequency Words

Prominent words are those that in a particular year appeared with a noticeably high relative frequency compared to earlier years. Likewise, a word that did not appear in earlier years and appears for the first time in a particular year, even only a small number of times, is also given a high value on the log-likelihood calculation (Rayson, 2008). It is impractical to comprehensively show the data provided by these calculations and we have thus chosen to present two interesting cases that emerged from them (Figures 6 and 7).

Figure 6 shows the change over time in the appearance of the word "Facebook". The data show that in 2011 this word is highly prominent, after not being mentioned at all in the years 2006-2010 in the Chais articles. Starting in 2011, it continues to be very prominent each year. The absence of the word "Facebook" from 2006 to 2010 and its spread from 2011 onward (despite the fact that Facebook started being used by the public already in 2006) can be explained by the fact that only in 2009 did it become possible to define languages other than English (including Hebrew) for use with personal profiles.
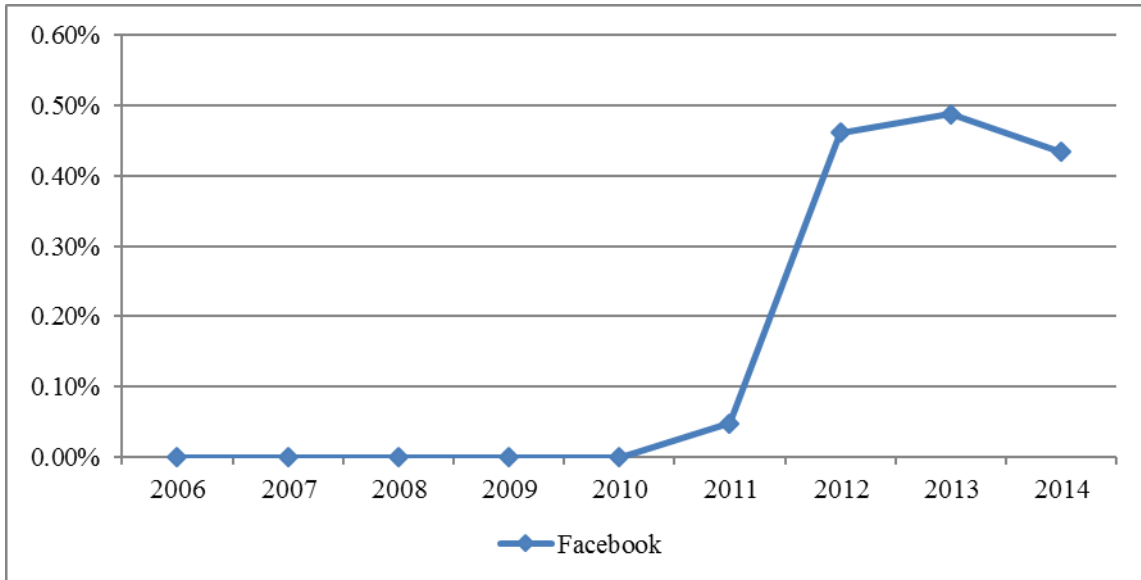
Figure 6. Development of the use of the word "Facebook" over the years

Figure 7 shows the five most prominent words in 2013: "classroom", "ICT", "program", "education", and "lesson".

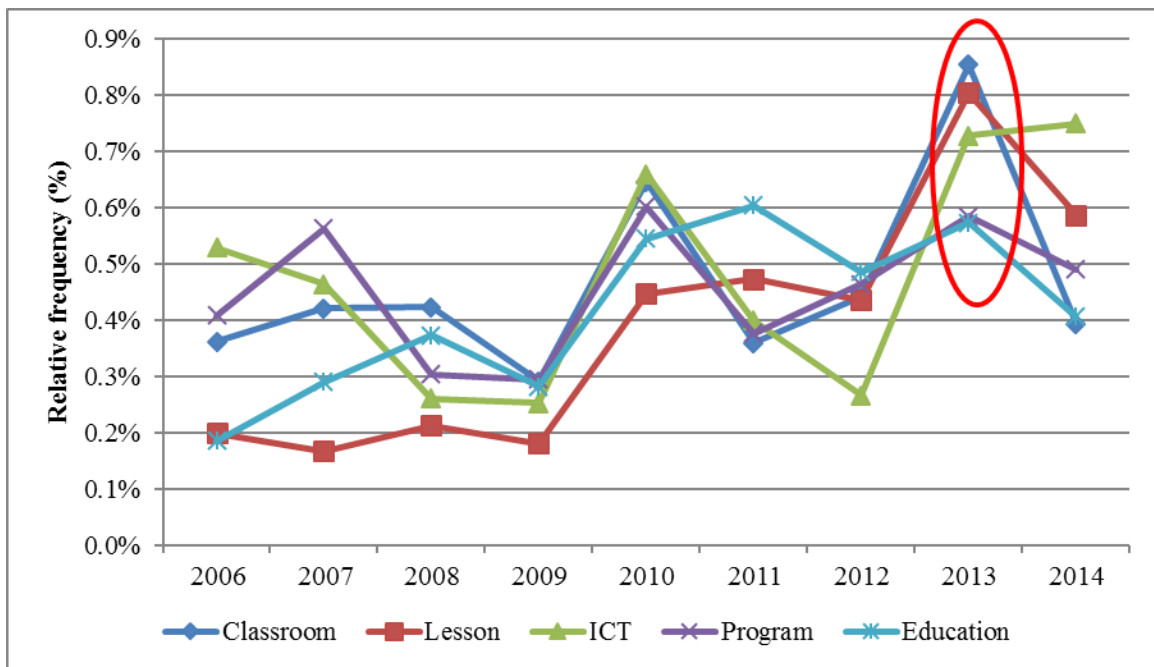**Figure 7. Five most prominent words in 2013**

Discussion

This study applied culturomics, a new data-driven approach of research, to identify characteristics and trends of change over time in research discourse in the field of learning technologies in Israel, as reflected in the research articles presented at the Chais Conferences for the Study of Innovation and Learning Technologies over a period of nine years.

Considering the research questions about the patterns of change over time in learning technologies research, the findings demonstrate the inherent potential of a data-driven approach for:

- Identifying common terms and expressions in the domain of learning technologies (such as "collaborative learning", "teachers", "students", "face-to-face").

- Identifying a prominent trend of research for a particular year and linking of the technology to the date of its first appearance (e.g., "Facebook", "the National ICT program").

- Identifying research trends: the most common terms belong to the domain of pedagogy.

An exhaustive process of analysis of the data for all words in the corpus unexpectedly revealed that the most common terms in this research field (e.g., "pupil", "teacher", "student") and the most common phrases (e.g., "face-to-face", "collaborative learning") are actually from the domain of pedagogy and not from the domain of technology. Thus, the picture that emerges from the current study is that this is not an area of research that actually focuses on technology, but rather on the pedagogical aspects of integrating technologies into teaching.

We analyzed two cases that demonstrate patterns of change within the prominent themes. First, the case of Facebook, which since 2011, emerged as a prominent trend of research (Figure 6). We conjecture that the option to define other languages in that platform significantly expanded the use of this social network, particularly among youth, which then led to research of "Facebook" with regard to learning technologies. Thus, the culturomics methodology managed to link the tool to the date of its first actual use in Israel. Second, we located the peak of "the National ICT program" and related terms in 2013 (Figure 7). Here, too, we can explain the findings obtained from the quantitative analysis as reflecting the many studies, which begun in 2012, following the launch of the National ICT program in Israel – "Adapting the Education System to the 21st century" (Ministry of Education, 2011), bringing the issue to the forefront of academic-educational debate in Israel. These findings demonstrate the potential of the data-driven approach for identifying trends of change in the field of learning technologies.

The findings of this study demonstrate the potential contribution of the data-driven approach to a retrospective understanding of the development of the field and the changes it is undergoing. With this approach we have presented an overview that constitutes the basis for future exploration of the changes over time in the use of terms and terminology, and such data can shed light on the "birth and death" of the noteworthy areas of interest in the field. This approach concurs with the trend of the growing use of data-driven analysis research in the era of Big Data (Michel et al., 2011; Radinsky et al., 2011; Radinsky & Horvitz, 2013).

Research Limitations and Directions for Further Research

In the current study, we showed that the corpus of academic articles with a scope of some 700,000 words had a Zipf-type frequency distribution, so that only 39% of the content appeared in a sufficient quantity to warrant processing. The relatively small number of words, their limited frequency and the range of only nine years hampered the analysis of discourse trends. It seems that the scope of the study's dataset (the amount of text and the span of years covered) is not sufficient for drawing conclusions on trends, which based on a quantitative analysis. In a study by Silber-Varod et al. (2016), the authors propose a mixed methods model for big data discourse analysis in which culturomics analyses should involve both quantitative and qualitative methods. In studies in which the quantity of text or the span of years is relatively small, the main research method should be qualitative. Conversely, the broader the scope of the dataset, the more one can rely on the quantitative analysis (Silber-Varod et al., 2016).

Moreover, one should remember that the frequency of the use of words is influenced by the style of writing, by new terminology in the field, or by words belonging to a particular jargon. These

factors might be the cause of the great variety in the choice of keywords the authors of the articles made, however we cannot learn about the reason for their use in this kind of study, and the interpretation of the findings requires the use of qualitative research methods or additional quantitative research tools. Further research might use language processing tools for semantic networks beyond the morphological analysis conducted in this study, since these tools may neutralize the influence of writing style or the use of synonyms.

In addition, the data we presented might have value as infrastructure data for studies that focus on a specific aspect such as "social network" and the technologies connected to it. Focused research questions such as these and others might benefit from the quantitative data generated in the current study. For this purpose, we have made the database of the 2006-2014 Chais conference articles available to discourse researchers on the Corpus Query Processor (CQP) (Hardie, 2012), which enables effective language and linguistic searching. This corpus can be found at: http://yeda.cs.technion.ac.il/HebrewCqpWeb/learningman/.

Conclusion

In the current study, several natural language processing tools were used to analyze the features and trends in the discourse of the Chais Conference articles pertaining to learning technologies research in Israel. The study shows that despite the relatively limited scope of the database, the use of relative frequency calculations and comparison of the prominence of lexical items over the years revealed trends and features in discourse on the subject. The main findings indicated the fundamental lexical items in the research discourse on learning technologies in Israel and showed that the most commonly used terms (e.g., pupil, teacher, student) and the most commonly used phrases (e.g., face-to-face) in this field reflect a pedagogical rather than a technological aspect of learning technologies. Furthermore, the study identified the most evident fluctuations in technological terminology across time and their association with what is taking place in the Israeli and global learning technologies arena, such as the rise of Facebook as a research area.

This research may contribute to developing retrospective understanding on how the research field of learning technologies has evolved over the past ten years in Israel. We showed data-driven insights on the pattern of change through time and the lifespan of research topics.

Methodologically, this research is in line with discourse data-driven research, which is an emerging trend in the Big Data era. To conclude, we suggest that expert reports could be strengthened by quantitative analysis of the textual database. Thus, this research also contributes to demonstrating the effectiveness of discourse analysis as a methodology for identifying evolving trends in fast-changing fields.

Acknowledgements

References

Anthony, L. (2011). Antconc version 3.2.4w. Center for English Language Education, Waseda University. Retrieved from: http://www.antlab.sci.waseda.ac.jp/software.html

Bakia, M., Murphy, R., Anderson, K., & Trinidad, G. E., (2011). International experiences with technology in education: final report. U.S. Department of Education, Office of Educational Technology, and the Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service.

Bennet, R. E., & Ben-Simon, A. (2005). Toward theoretically meaningful automated essay scoring. Report to the National Institute for Testing and Evaluation (NITE). Retrieved from: https://nite.org.il/files/reports/e329.pdf

Ben-Simon, A., & Cohen, Y. (2011). The Hebrew Language Project: Automated essay scoring & readability analysis. *The 37th International Association for Educational Assessment (IAEA) Annual Conference*, Manila. Retrieved from: http://www.iaea.info/documents/paper_4e1237ae.pdf

Ben-Simon, A., & Safran, Y. (2012). The factorial structure of written Hebrew and its application to AES. Paper presented at *The annual meeting of the National Council of Measurement in Education (NCME)*. April 12-16 2012, Vancouver, Canada. Retrieved from: https://www.nite.org.il/files/reports/e382.pdf

Bohannon, J. (2011). Google Books, Wikipedia, and the future of Culturomics. *Science*, *331*(6014), 135. 14 January 2011. Retrieved from: http://www.terceracultura.net/tc/wp-content/uploads/2011/01/culturomics.pdf

Cuban, L. (1999). The technology puzzle. *Education Week*, *18*(43), 68-69.

Cuban, L., Kirkpatrick, H., & Peck, C. (2001). High access and low use of technologies in high school classrooms: Explaining an apparent paradox. *American Educational Research Journal*, *38*(4), 813-834.

Culp, K. M., Honey, M., & Mandinach, E. (2005). A retrospective on twenty years of education technology policy. *Journal of Educational Computing Research*, *32*(3), 279-307.

Elgali, Z., & Kalman, Y. M. (2011). Effects of ICT fashions on national educational technology policy. Paper presented at *the 6th Mediterranean Conference on Information Systems (MCIS)*, Limassol, Cyprus, September 3-5, 2011.

Eshet-Alkalai, Y., & Chajut, E. (2009). Changes over time in digital literacy. *Cyber Psychology & Behavior*, *12*(6): 713-715. doi:10.1089/cpb.2008.0264.

Geri, N, Blau, I., Caspi, A., Kalman, Y. M., Silber-Varod, V., & Eshet-Alkalai, Y. (2015). A decade of Chais conferences: Introduction to the IJELL special series of Chais conference 2015 best papers. *Interdisciplinary Journal of e-Skills and Life Long Learning*, *11*, 149-157

Johnstone, B. (2008). *Discourse analysis* (2nd ed.). Malden, MA: Blackwell Publishing.

Hardie, A. (2012). CQPweb – Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, *17*(3), 380-409.

Leydesdorff, L., & Milojevič, S. (2015). Scientometrics. In J. D. Wright (editor-in-chief), *International Encyclopedia of Social and Behavioral Sciences* (2nd ed.), Section 8.5: Science and Technology Studies, Subsection 85030. Micheal Lynch (ed.), Vol. 21, pp. 322-327. Oxford: Elsevier

Livnat, Z. (2012). *Dialogue, science and academic writing* (Vol. 13). John Benjamins Publishing.

Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., ... & Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176-182.

Ministry of Education. (2011). Adapting the educational system to the 21st century [in Hebrew]. (Master Document. Version 11 31-08-2011). Retrieved from http://cms.education.gov.il/NR/rdonlyres/79B5A8CF-F812-4A63-89BE-3BEFEB887EC5/142454/12.pdf

Raban, D. R., & Gordon, A. (2015). The effect of technology on learning research trends: A bibliometric analysis over five decades. *Scientometrics*, *105*(1), 665-681.

Radinsky, K., Agichtein, E., Gabrilovich, E., & Markovitch, S. (2011). A word at a time: Computing word relatedness using temporal semantic analysis. In *Proceedings of the 20th International Conference on the World Wide Web* (pp. 337-346). ACM. Retrieved from http://www.gabrilovich.com/publications/papers/Radinsky2011WTS.pdf

Radinsky, K., & Horvitz, E. (2013). Mining the web to predict future events. In *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining* (pp. 255-264). ACM.

Rayson, P. (2003). *Matrix: A statistical method and software tool for linguistic analysis through corpus comparison* (Doctoral dissertation, Lancaster University).

Rayson, P. (2008). From key words to key semantic domains. *International Journal of Corpus Linguistics*, *13*(4), 519-549.

Rayson, P. & Garside, R. (2000). Comparing corpora using frequency profiling. In *Proceedings of the Workshop on Comparing Corpora, held in conjunction with the 38th Annual Meeting of the Association for Computational Linguistics (ACL 2000)*. 1-8 October 2000, Hong Kong, pp. 1-6.

Silber-Varod, V., Eshet-Alkalai, Y., & Geri, N. (2016). Culturomics: Reflections on the potential of big data discourse analysis methods for identifying research trends. *Online Journal of Applied Knowledge Management*, *4*(1), 82-98.

Sinclair, J. (1991). *Corpus, concordance, collocation* (Vol. 1). Oxford: Oxford University Press.

Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Oxford: Routledge.

Soper, D. S., & Turel, O. (2012a). Who are we? Mining institutional identities using n-grams. *System Science (HICSS), 2012 45th Hawaii International Conference on*, 1107-1116. IEEE.

Soper, D. S., & Turel, O. (2012b). An n-gram analysis of communications 2000-2010. *Communications of the ACM*, *55*(5), 81-87.

Soper, D. S., Turel, O., & Geri, N. (2014). The intellectual core of the IS Field: A systematic exploration of theories in our top journals. In *System Sciences (HICSS), 2014 47th Hawaii International Conference on* (pp. 4629-4638). IEEE.

Van den Besselaar, P., & Heimeriks, G. (2006). Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics*, *68*(3), 377-393.

Appendix A: The Database

Eshet-Alkalai, Y., Caspi, A., Geri, N., Kalman, Y., Silber-Varod, V., & Yair, Y. (Eds.). (2014). *Learning in the technological era: Proceedings of the 9th Chais Conference for the Study of Innovation and Learning Technologies*, Raanana, Israel: Open University Press, 455 pp. (partly Hebrew). Available at: http://www.openu.ac.il/innovation/chais2014/program.html

Eshet-Alkalai, Y., Caspi, A., Eden, S., Geri, N., Yair, Y., & Kalman, Y. (Eds.). (2013). *Learning in the technological era, Proceedings of the 8th Chais Conference for Innovation in Learning Technologies Research*, Raanana, Israel: Open University Press, 454 pp. (partly Hebrew). Available at: http://www.openu.ac.il/innovation/chais2013/papers.html

Eshet-Alkalai, Y., Caspi, A., Eden, S., Geri, N., Yair, Y., & Kalman, Y. (Eds.). (2012). *Learning in the technological era, Proceedings of the 7th Chais Conference for Innovation in Learning Technologies Research*, Raanana, Israel: Open University Press, 441 pp. (partly Hebrew). Available at: http://www.openu.ac.il/innovation/chais2012/papers.html

Eshet-Alkalai, Y., Caspi, A., Eden, S., Geri, N., & Yair, Y. (Eds.). (2011). *Learning in the technological era, Proceedings of the 6th Chais conference on instructional technologies research*, Raanana, Israel: Open University Press, 393 pp. (partly Hebrew). Available at: http://www.openu.ac.il/research_center/chais2011/papers.html

Eshet-Alkalai, Y., Caspi, A., Eden, S., Geri, N., & Yair, Y. (Eds.). (2010). *Learning in the technological era, Proceedings of the 5th Chais conference on instructional technologies research*, Raanana, Israel: Open University Press, 381 pp. (partly Hebrew). Available at: http://telem.openu.ac.il/hp_files/chais/10/site.html

Eshet-Alkalai, Y., Caspi, A., Eden, S., Geri, N., & Yair, Y. (Eds.). (2009). *Learning in the technological era, Proceedings of the 4th Chais conference on instructional technologies research*, Raanana, Israel: Open University Press, 421 pp. (partly Hebrew). Available at: http://telem.openu.ac.il/hp_files/chais/09/articles.html

Eshet-Alkalai, Y., Caspi, A., & Geri, N. (Eds.). (2008). *Learning in the technological era, Proceedings of the 3rd Chais conference on instructional technologies research*, Raanana, Israel: Open University Press, 396 pp. (partly Hebrew). http://telem.openu.ac.il/hp_files/chais/08/articles.html

Eshet-Alkalai, Y., Caspi, A., & Yair, Y. (Eds.). (2007). *Learning in the technological era, Proceedings of the 2nd Chais conference on instructional technologies research*, Raanana, Israel: Open University Press, 382 pp. (partly Hebrew). Available at: http://telem.openu.ac.il/hp_files/chais/07/site.html

Eshet-Alkalai, Y., Caspi, A., & Yair, Y. (Eds.). (2006). *Learning in the technological era, Proceedings of the 1st Chais conference on instructional technologies research*, Raanana, Israel: Open University Press, 291 pp. (partly Hebrew). Available at: http://www.openu.ac.il/innovation/chais2006/

Biographies

Vered Silber-Varod is a Research Fellow at The Research Center for the Study of Innovation in Learning Technologies, The Open University of Israel. She holds a B.A. in Political Science and French language and literature studies from the Hebrew University of Jerusalem, an M.A. with Magna Com Lauda in Hebrew language studies at the department of Hebrew and Semitic languages from Tel Aviv University, and a Ph.D. in Humanities from Tel Aviv University. Her research interests and publications focus on various aspects of linguistics, with expertise in speech prosody, acoustic phonetics, and quantitative analysis of written and spoken texts. Vered has over 10 years of experience as a Linguist at various institutes, including the ACLP – Afeka Center for Language Processing, Afeka College of Engineering. Personal site: http://www.openu.ac.il/en/personalsites/VeredSilberVarod.aspx

Yoram Eshet-Alkalai is a Professor at the Open University of Israel, Department of Education & Psychology. He is Head of the M.A. program in educational technology, and the former Head of the Research Center for Innovation in Learning Technologies in the Open University. He holds a B.A. in Archeology (Hebrew University), M.Sc. in Geology (Hebrew University) and Ph.D. in Earth & Environmental Sciences (City University of NY). For a decade, he was the Head of the Instructional Design Program in the Tel Hai Academic College, and a senior researcher in the Geological Survey of Israel. He has 15 years of experience in developing technology-based instructional solutions for educational systems in Israel and the USA. In this capacity, he was involved in the design of hundreds of simulations, data-bases, tutorials and large-scale curriculum integration projects. His major research and publications interests focus on cognitive aspects of working with digital technologies, digital games and design principles of computer-based learning environments. Personal site: http://www.openu.ac.il/Personal_sites/yoram-eshet.html

Nitza Geri is an Associate Professor at the Open University of Israel, Department of Management and Economics, and Head of the Research Center for Innovation in Learning Technologies. She holds a B.A. in Accounting and Economics, an M.Sc. in Management Sciences, and a Ph.D. in Technology and Information Systems Management from Tel-Aviv University. Nitza is a CPA (Israel) and prior to her academic career she had over 12 years of business experience. Her research interests and publications focus on various aspects of the value of information and knowledge, as well as adoption, implementation, and continued use of information systems, including strategic information systems, e-business, information economics, attention economy, knowledge management, value creation and the Theory of Constraints, managerial aspects of e-learning systems. Personal site: http://www.openu.ac.il/Personal_sites/nitza-geri.html