

Article Illustration of a Software-Aided Content Analysis Methodology Applied to Educational Research

Maria Gkevrou and Dimitrios Stamovlasis *D

Department of Philosophy and Education, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece; marigkev@edlit.auth.gr

* Correspondence: stadi@edlit.auth.gr

Abstract: This paper presents a software-aided methodology for content analysis by implementing the Leximancer software package, which can convert plain texts into conceptual networks that show how the prevalent concepts are linked with each other. The generated concept maps are associative networks of meaning related to the topics elaborated in the analyzed documents and reflect the creators' core mental representations. The applicability of Leximancer is demonstrated in an education research context, probing university students' epistemological beliefs, where a qualitative semantic analysis could be applied by inspecting and interpreting the portrayed relationships among concepts. In addition, concept-map-generating matrices, ensuing from the previous step, are introduced to another specialized software, Gephi, and further network analysis is performed using quantitative measures of centrality, such as degree, betweenness and closeness. Besides illustrating the method of this semantic analysis of textual data and deliberating the advances of digital innovations, the paper discusses theoretical issues underpinning the network analysis, which are related to the complexity theory framework, while building bridges between qualitative and quantitative traditional approaches in educational research.

Keywords: textual data; content analysis; concept maps; Leximancer software; network analysis; complexity; epistemology

1. Introduction

The digital era of the 21st century has increased the demand for processing enormous quantities of data and specifically text-type data, which is freely accessible and available in the rapidly growing literature. In this framework, a variety of sophisticated methods have been developed for skim reading when perusing reviews or automatic text synopses. These are readily utilized in social and education sciences as means of facilitating practice and enhancing the effectiveness of relevant processes. In addition, they consist of valuable tools in research when adapted to specific inquires and comprise the state-of the-art methodologies for treating empirical data. Such an example is the content or semantic analysis of texts that is grounded on statistical processing. The techniques applied to content analysis vary with the basic philosophy fostered, which is closely related to the algorithm that is applied by the corresponding software, while different names were given, such as, hyperspace analogue to language [1], latent semantic analysis [2] or Leximancer [3]. which is an advanced natural language processing software utilizing Bayesian theory.

A central aspect is the notion of a *concept* and its visual representation, which is known as a *concept map* that depicts relationships as connections among other concepts. In the case of texts, concept maps are derived algorithmically by employing techniques such as correspondence analysis [4] and self-organizing maps [5], while other approaches follow specifically developed tools, namely Tetralogie [6], Terminoweb [7] or Leximancer [3], to mention a few.



Citation: Gkevrou, M.; Stamovlasis, D. Illustration of a Software-Aided Content Analysis Methodology Applied to Educational Research. *Educ. Sci.* **2022**, *12*, 328. https://doi.org/10.3390/educsci12050328

Academic Editor: Ana Garcia-Valcarcel Munoz-Repiso

Received: 8 April 2022 Accepted: 4 May 2022 Published: 7 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). These techniques are well-established in the domain of text mining, and the extraction of concept maps from texts in a visualized form has been a determining step for content analysis in both qualitative and quantitative approaches.

The question, however, that is of central interest in these endeavors is whether and how concept maps can be used to interpret and understand the overall structure of the data source, and to enable exploring the detailed content of the text corpus. The contemporary digital innovations in methodology focus on answering this question and propose means for appraising a concept within its global context. An interesting and effective solution is offered by the Leximancer software.

The Leximancer has previously been used to explore a multitude of research areas, such as analyzing polls and political commentary [8], investigating the communication strategies used by providers of care for people with schizophrenia [9], tracking the history of research articles in the Journal of Cross-Cultural Psychology to identify potential shifts in research interest [10], phenomenographic analysis of students' perceptions of literacy [11], etc. It is apparent, therefore, that the use of this tool is a potentially valuable choice in many areas of inquiry.

The present paper illustrates the applicability of software-aided content analysis in educational research by using the Leximancer package. It demonstrates that the main or dominant concepts at a global level can be revealed via a visual representation that allows the probing of local relationships [12]. The method is described via a thorough data analysis taken from an educational enquiry, while the basic methodological aspects regarding the underlying algorithms are explained. In addition, theoretical and epistemological issues underpinning this application are discussed, as this interdisciplinary approach lies at the interface between the traditional qualitative and quantitative methodologies.

2. Semantic Representations

The methodology presented in this paper is based on certain ontological considerations regarding the subject under study, that is, human perceptions, knowledge or beliefs, and any latent variable is viewed as an associative network of concepts. When investigating a text corpus, the conceptual space of interest can also be represented as an associative network of meaning that encompasses core and peripheral concepts, portrayed in a *concept map*, which refers to a pictorial illustration showing how these elements are linked to each other [13]. More specifically, concept maps are web-type representations that depict the kind and the strength of relationships through semantic or meaning-driven connections [14]. In short, the starting point and the development of concept maps is situated in the need to represent conceptual understanding and knowledge [15], providing an illustrative portrayal of an individual's internal conceptual structure [14]. The methods of forming concept maps are used in both research and educational processes and are suitable for individual and collective projects in a knowledge domain, while for constructing them directly by the subjects, suitable software have been developed (e.g., CmapTools) [16,17]. Certainly, the process of forming a concept map is rigorous, because the transfer of important information at the level of understanding is particularly demanding [18]. It is remarkable that digital innovations offer the possibility to address this issue via a software-aided natural language process that reveals concept maps from text corpora, fostering the above-mentioned ontology of conceptual representations.

A concept can be a single theoretical category, which can be expressed in a single word, a nominal phrase or even a verbal phrase [19]. But within its broader definition a concept is an ideal, abstract core that is concretized in its association with other concepts [20]. A relationship, then, refers to the bond that connects two concepts [19,20]. This relation, however, is not univocal; instead, it may express conceptual relationship, or it may indicate that the concepts are somehow cognitively related in the mind of the author of the text, or it may simply express their proximity in the text [19], and at the same time they may possess directionality, strength, sign, and meaning [20]. The relationship, then, between two concepts (nodes in the network) can be one- or two-way path, while the strength of

this relationship can also vary. Furthermore, the sign of a relation can be distinguished by indicating either a positive or a negative connotation, and of course the relation between two concepts can differ in meaning.

Moreover, a statement consists of two concepts and the relationship between them, while the map as a network of concepts is formed by more than one statement. Two statements are connected if they share a concept, and these statements can start forming a network, or otherwise, a map. Managing and supervision of their aggregation process are achieved by software algorithms designed to covert the textual data to a network of meaning.

A map extracted from the encoding of a text created by an individual can be seen as a representation of her/his mental model. Furthermore, a map created by joining two individual maps can be seen as a group map, or a representation of shared mental models [21]. Participants in concept map creation work individually when providing their ideas and their unstructured or yet unstructured beliefs, and the collective group perception emerges from the analysis of the overall individual data. Collective concept maps can also be derived when analyzing text segments together, that is, when a group of individuals have provided different inputs on a particular topic. They actually represent the shared features that are extracted without requiring any consensus among them. It is imperative to emphasize that the use of collective maps is an advantageous approach because the joined network of meaning has been derived with no contact or interactions between individuals, and the negative effects of group dynamics in a discourse process that can allow diverse thinking to emerge are definitely avoided [21].

In conclusion, semantic representations originating from content analysis [22] by natural language processing with the contribution of network theory and other scientific domains, reveal the cognitive edifices and mental models laying beneath the original data, exemplifying their complex structure and properties.

3. Complexity

It is pertinent to explicate that the ontology of semantic networks belongs to the metatheoretical framework of *complexity science*. Complexity theory assumes the ontological features that are described by *networks*, that is, many interconnected and interacting parts co-evolving in time. Even though time is implicit in the traditional endeavors involving text analysis, the relationships between concepts are considered as having been dynamically formed from which the global features of the whole network emerge, and this is a manifestation of complexity [23,24]. In addition, network science provides the mathematical formulation to study those complex systems and explore the internal structure consisting of nodes and links. In the domain of cognition (scientific or ordinary), personal conceptualizations can be described as a web, where concepts, principles and other types of conceptual elements are connected in a complex way, forming a knowledge system. A basic aspect of a complex system is that its properties and behavior are described in terms of the underlying interacting components as well as in terms of the system as a unit or a whole. In such a system, the notions of quality and quantity are no longer distinct features. They characterize the same system depending on the level of complexity that is observed, the micro or macro level, which is examined within a unified framework [25–27]. Thus, the present methodology attempts to connect the two levels of complexity, that is, what the designers of the data mining tools intend when exploring text-corpus data of a particular domain and focusing on the high-level/macro view and on the specific-details/micro level, as well [12].

The complexity paradigm has fostered numerous works in social science, where common networks of physical entities comprise relational links with connections that can represent friendship, relationship, cooperation, actions or interaction [28–30]. On the other hand, there are immaterial webs with non-physical links that can be either semantic or linguistic networks, and these are the interest of the present work. In recent years, network analysis has been used to identify social and cognitive patterns in a variety of situations,

including educational contexts. Further advances of network analysis are implemented to understand associative networks of meaning on different areas, i.e., students' ideas of complex systems, knowledge of history of sciences, classroom processes or evaluation of complex educational systems [31–34], to mention a few. Relevant to the present inquiry, an intriguing application of network analysis, in education research, is on knowledge-believe domain, where internal cognitive structures are reflected and approximated on students' concept maps [13].

It is important that the methodological approach examined in the present paper is not merely a guileless choice offered by digital invocations, but is a theory-supported framework where technology is just a blessed facilitator.

4. Content Analysis with the Leximancer Software Package

The Leximancer is a software package developed for performing conceptual analysis of natural language textual data and consist an unsupervised approach that uses master concept classifiers [35,36]. It actually incorporates a content analysis method with advanced analytic techniques that have been adopted from the cutting-edge grounds, such as information science, network theory, computational linguistics and machine learning [37]. The Leximancer software performs a semi-automatic analysis to a text type of data, from where it extracts core concepts and based on their co-occurrence, it generates a concept map. The basic features and steps of the method are as follows.

The Leximancer mapping procedures are operating in two sequential stages, which are characterized as the semantic and the relational extraction, respectively [38], where both use data of sporadic co-occurrence records. In the semantic-extraction step, the algorithm used, process term-by-term, small segments of the document, i.e., one to three sentence, at the extent of a paragraph or more. The semantic representation that is achieved, even with small datasets, implements Bayesian co-occurrence metrics [12], which are efficiently used in performing text classification. In the next step, the relationship-extraction, stage, the learned concept classifiers are used to categorize the text segments, while a frequency table of concept co-occurrence is computed. The diagonal elements of this matrix representing the co-occurrence frequency, serve to normalize the co-occurrence in each column expressing the corresponding probability. The table with the selected concepts and along with their conditional probabilities, comprise the input information for the algorithm constructing the ultimate mapping [12,36]. The computation produces a set of encoded locations in a surface/overlapping space geometry, where distances between concept-nodes is the degree of their relevance, that is, smaller distances represent higher associations.

In summary, the Leximancer program, from an integrated body of text, creates an ordered list of the most important lexical terms, where frequencies and co-occurrence metrics are the ranking criteria. The collected words and terms are then used to build a thesaurus that provides a set of classifiers which are named as concepts [38]. These concept tags/encodings are registered to provide a document exploration environment for the user. Then, based on these classifiers, the text is examined at a high-resolution level, and it is categorized by multi-sentence analysis. The procedure results in an index of concepts and their co-occurrence matrix.

The number of concepts and their recorded co-occurrences determine the complexity of their relationships, the illustration of which needs advanced algorithmic procedures. Complex systems approaches have been applied, and several new algorithms have been developed, such as learning optimizers to automatically select, learn and adapt a concept from word usage within the text and an asymmetric scaling process to create a conceptual cluster map based on co-occurrence in the text [35]. Thus, the matrix of co-occurrences is implemented to construct the 2D conceptual map by means of an appropriate clustering procedure. The connectivity of each node concept in the ensued semantic web creates a hierarchical structure as well, which could display more general nodal groups at a higher level of complexity [38].

The principal objective of the Leximancer philosophy is to reveal and enable the researcher to understand the global structure of the text under study. There are featured concepts that dominate and characterize the mental representations emerging from the text which convey the main message, and they should be acknowledged. On the contrary, incidental and insignificant evidence, which could be of minor meaning or random atypical references, should be ignored.

Besides knowing the ultimate goal of the Leximancer-based content analysis, the researcher is also facilitated in the inquiry process by learning how the algorithms are applied and what choices they made in the course of the text analysis, and what the interpretation of the graphical and representational feature of the outputs is. In Leximancer system analysis, the frequently occurring words are preserved as concepts. The supplied output that includes the sought network providing the conceptual synopsis of the initial textual corpus demands an interpretation. This is facilitated by observing larger groups/clusters assembled by topic circles and names as *themes* that include concepts which encapsulate the same main idea and are named after the supreme concept in the cluster [9]. Note that theme/central idea is not synonymous with the term 'concept', but themes are clusters of concepts. On the display, these are heat-mapped, meaning that warm colors correspond to the most relevant themes and cool colors to the least [39]. Based on the produced concept map and the lists/thesauruses of classified concepts, the observer can capture and understand the relationships among concepts. The visual depiction highlights the strength of the associations and provides a global overview of the semantic network representing original textual data, while color coding facilitates the distinction between salient and less important node concepts [9,40].

The development of a qualitative analysis software packages, such as Leximancer, facilitates and automates the coding process, with speed and consistency, particularly when a sizeable amount of information is collected and transcribed [36]. Note that, whereas there is no a unique way to analyze texts, the design of Leximancer provides an holistic viewpoint of themes and central ideas, and attempts to enhance objectivity and avoid biases, comparing to the traditional manually-performed content analysis [38]. Last, an advantage of the program is that it allows the use of many languages, and this is important component for the international researcher.

Other Leximancer Outputs and Perspectives

In the previous sections, concept maps as semantic representations ensuing from the analysis process via Leximancer were discussed in detail. In this section, other output capabilities of the software that offer even deeper insights into the data are presented without describing them thoroughly. In addition to concept maps as pictorial representations of textual data, the Leximancer can extract the so-called concept clouds, which also comprise a visualization tool. Concept clouds, as networks, depict the most frequent and relevant concepts extracted from the data. As with concept maps, more relevant concepts are presented in warm colors (red, orange) and less relevant concepts in cooler colors (blue, green), while the size of the node denoting a concept corresponds to and measures the frequency of that concept and the relativeness between two or more concepts, is determined by the distance between them [41].

Another feature that Leximancer offers for reviewing and understanding the data is the report Insight Dashboard, which provides a more quantitative approach to the data as well as a graphical representation. This application facilitates a more general overview of the data files under consideration. The graphical representation is composed of quadrants in which strength (vertical) and relative frequency (horizontal) are plotted. This information is compiled in tables as reference tools which are offered for comparison or as an analysis of differences [39]. The statistical tables also reflect the prominence, which is obtained as a result of strength and frequency scores using Bayesian statistics [41]. The advantages of this application are its availability to other researchers and its comprehensibility without any prior knowledge of the software [41]. Finally, upon completion of the data processing, the program generates statistical results, which can be imported into other network analysis software for further examination, such as the program *Gephi*, where quantitative characteristics of the ensued network, such as closeness centrality, betweenness centrality, average degree of network, etc. could be estimated. In summary, Leximancer offers: a matrix of co-occurrence of concepts across the text, the list or pairs of concepts in each text excerpt, the list of concepts in each context block and the seed set of the sentiment lens, if enabled [41].

The possibility of supportive cooperation with other popular programs for additional investigations of the ensuing semantic maps gives a perspective of promising applications.

5. An application in Educational Research

5.1. Probing Students Epistemological Believes: An Educational Inquiry

In education and generally in social sciences, theoretical and methodological aspects are highly affected by the epistemological orientations that researchers follow. The relevant rich literature is illuminating about the variety of approaches and perspectives existing within social science community and the historical process of their genesis, the deeper explicit differences or the common aspects are well known. This endeavor focusses on the two general and dominant epistemological trends: namely: the epistemological positions of positivism and constructivism respectively. The latter has been identified with both hermeneutic and critical perspectives, while a third one subsists, the stance of skepticism, that has been derived from the postmodern view [42–46], which however will not be considered here. The two main philosophical positions correspond to distinct and well recognizable frameworks, from which different methodological approaches in social science research are derived.

The present work belongs to a broader research endeavor on philosophical underpinnings of research methodology and focusing on personal epistemologies, explores university students' epistemic beliefs in conjunction with the formal views, that is, by taking into consideration the positions of positivism and constructivism as they are stated in the literature [42–46].

Research on students' epistemological perceptions and beliefs has shown that they play an essential role, because they affect students' attitudes toward science and scientific knowledge. They also contour many relevant individual differences and academic behaviors [47] such as motivation, achievement goal orientations or learning strategies [48]. The literature investigating students' epistemological beliefs is rich, and it has been accomplished via traditional qualitative and quantitative methodologies. The former, with an exploratory mode, have attempted to identify elements of knowledge and formal philosophical orientations based on the explicit responses of participants, while the latter were used to establish correlations between the epistemic views and other individual differences.

The two epistemic 'schools of thought' in question appear as the dominant orientations in the majority of methodology textbooks, along with a plethora of comparable variations in alternative philosophical stances. These lead to miscellaneous inquiry schemes and obviously ascertain that in social sciences, a fragmented epistemological and methodological landscape exists [25], which has an anticipated impact on students' attitudes and personal epistemic views. Scholars involved in teaching research methodology are mindful that the above impose difficulties on students and novice researchers, who have to choose their approach to empirical research between alternatives with unclear criteria,. The epistemological framework that students finally foster is being molded during their college and everyday life, influenced by academic and non-academic sources. This characterizes social science and education research and eventually permeates the academic lecture halls through the curricula and the texts of the researcher-teachers who choose one or the other epistemological framework, matters that induce difficulties and confusion among students. Thus, investigating students' personal epistemologies provides an enlightened picture of the education system's impact on the participant-students, who constitute the future scientific community and possibly the research academic workforces.

The present endeavor aims to explore students' epistemological beliefs by fostering the newest theories of latent variable representation and implementing content analysis based on *semantic mapping*. According to this methodology, the epistemological beliefs and their element concepts are not restricted to their lexical definitions but are represented and studied as associative webs of meaning [48]. As psychological constructs, epistemological beliefs are understood and represented as networks of elementary concepts connected in various ways, forming an emergent qualitative entity. This approach is in line with contemporary psychometric theories and analyses [49,50], but the sought networks are constructed in different way, while a qualitative or a quantitative evaluation could be applied. Specifically, in this endeavor, the networks in question emerge from content analysis carried out via a computer-aided process, offered by the Leximancer software package.

The present analysis focused on textual data derived from the participation of 110 undergraduate students enrolled in the social sciences and humanities. The total collected material examined here amounts to about 180 pages of text. The participants were attending a research methodology course, and they had to write and deliver essays which develop arguments for or against certain philosophical hypotheses about reality and the validity of potential knowledge about it. The research and data collection procedures adhered to the rules and were approved by the Institutional Ethics and Deontology Committee.

5.2. Analysis of Textual Data

The textual data were introduced to the Leximancer software program, and the outputs as conceptual networks were collected for further analysis and interpretations. Once the corpus of textual data has been entered into the program, the researcher, in the next stage of the analysis, takes the option of removing articles, linking words and stopping words and other word concepts that are deemed to add nothing meaningful semantically and conceptually to the formation of the concept maps [8]. This option is considered to be the process of removing 'noise' from the data, and it is necessary for attaining a meaningful representation of the concept maps.

Next, specific concept maps derived from empirical textual data are examined, which originate from students' essays elaborating on their perceptions and ontological beliefs about social reality, including subjectivity and/or objectivity issues.

Figure 1 shows the original Leximancer output result, while Figure 2 is a reconstructed and translated picture. Three main central themes that ensued from the analysis can be observed: "reality", "social" and "construction". The portrayed characterization is based on the size of the cluster that depends on both the individual embedded concepts, i.e., the concepts that the central and a localized relation coexists within the context, and on the repetition of the concepts. Table A1 in the Appendix A provides the quantitative indexes, percentages and conditional probabilities for this characterization. Thus, a collective concept map results from the statistical processing of the textual data, consisting of these central themes/clusters.

The strongest central theme, *reality*, includes the concepts: 'rules', 'people', 'interaction' and 'language'. The second strongest central theme, *social*, consists of the concepts: 'nature' and 'alteration' and in the third strong central theme, *construction*, the concepts 'objective' and 'subjective' are identified.

The concept map revealed that the foremost concepts in terms of occurrence frequency and relevance rates were: "reality", "social", "people", "interaction" and "construction". It seems, therefore, that these concepts can also act as attractors in clustering, as the quantitative indexes for these five prevailing concepts suggest (Table A1).



Figure 1. The emergent concept map from the analysis of texts elaborating the ontological question about social reality—Original Leximancer output.



Figure 2. The emergent concept map from the analysis of texts elaborating the ontological question about social reality—reconstructed translated picture from Figure 1.

As it was mentioned in a previous section, another pictorial output, the concept clouds, are provided by Leximancer, depicting the most frequent and relevant concepts extracted from the data. (Figure 3). Here, in terms of co-occurrences of the primary concepts, the concept "reality" coexists strongly with the concepts "social", "people", "interaction", "construction", etc. (Table A2 in the Appendix A). It seems that regardless of the grouping of concepts in a cluster, there are strong overlaps across the map. Similarly, the concept "social" co-exists strongly with "interaction", "construction" and "rules". The primary concept "people" co-exists strongly with "interaction", "construction" and "language". It is worth noting that as we move on to less strong concepts, the strength of the co-occurrences also decreases (see Table A2). Subsequently, the concept "interaction" is found to coexist with the concepts of "construction", "language" and "subjective". Finally, the concept "construction" seems to co-occur with "interaction", "subjective" and "objective".



Figure 3. The cloud view originating from the analysis of text elaborating on the question of the existence of social reality.

The above illustrates the process of semi-automatically extracting concept maps from social science students' reports elaborating on the nature of reality. While the findings are intriguing, the discussion will be limited since the scope of this paper is to illustrate the methodological issues. The ensuing concept maps can enable the comprehension of emerging primary ideas, as well as the underlying knowledge and belief structure of the text corpora creators. The main bound concepts with the most and the strongest cooccurring relationships appearing in this concept map are "reality", "social", "people", "interaction" and "construction". Another noteworthy finding in the concept map is the relationship depicted between the concepts "subjective" and "objective", with an almost negligible distance between them. The overall appraisal of the emergent semantic network leads to the conclusion that the map and the latent, collective philosophical belief system is leaning towards adopting the constructivist framework. This finding could help in future intervention design in the case of targeting a conceptual change or strengthening existing conceptual structures. It seems, therefore, that Leximancer can indeed provide a generalized overview of the textual data, giving useful insights for both inference and future educational practice in this domain.

5.3. Network Analysis of Concept Map Ensuing from Textual Data: Quantitative Measures

The previous section presented in detail the process and the results from the analysis of the textual data using the Leximancer program that gives an overall perspective of the findings and a primary understanding of the students' mental representations and their belief system under study. The preceding qualitative appraisal of the concept maps, as it has been mentioned earlier, can be followed by a formal network analysis [49].

Leximancer provides a suitable matrix which can be used as input for popular network analysis software, such as Gephi, and reproduces a detailed network which could be subjected to further analysis. In the present case, entering this data matrix into the Gephi software resulted in a concept network consisting of 11 nodes and 117 edges. Figure 4 depicts this complex web of concepts drawn by the Gephi program based on Leximancer output information. A network can be analyzed metrically at three levels: microscopic (single nodes), mesoscopic (clusters of nodes), and macroscopic (whole network) [50,51].



Figure 4. The network of concepts drawn by the Gephi program based on Leximancer output information.

The present study focuses the analysis at the microscopic level, i.e., at the node level, in order to identify the most important and central elements/nodes of the web. The most informative measures corresponding to this scale or level of complexity are the centrality measures, namely: degree, closeness centrality and betweenness. In short, the degree accounts for the number of direct links of a node, closeness centrality estimates the degree to which a node is related to other nodes in the network, and betweenness centrality estimates the degree to which a node is on the shortest path between two nodes [50]. Table 1 presents the measures calculated above for the network depicted in Figure 4.

Label	Closeness	Betweenness	Degree
reality	1.000	0.564	18
people	1.000	0.564	12
construction	1.000	0.564	12
rules	1.000	0.564	10
objective	0.928	0.182	10
subjective	0.867	0.000	16
interaction	1.000	0.564	16
alteration	0.867	0.000	8
nature	1.000	0.564	4
language	1.000	0.564	14
social	1.000	0.564	20

Table 1. Centrality measures calculated for the concept-nodes prevailing in the network ensuing from Leximancer (Gephi software output).

In relation to the measure of degree, the concepts with the most direct connections are: "social", with 20 links, "reality," with 18 connections, "subjective" and "interaction", both with 16 links, that is, they are strongly bonded with the network. Regarding the measure of closeness centrality, most of the concept nodes equally have a value of 1, except the concepts of "objective" with 0.928, "subjective" and "alteration" with 0.866. Closeness measures advocate that all these concepts are equally definitive of the core semantic network that reflects the epistemological belief system under investigation. As far as betweenness centrality is concerned, again, it shows that the majority of the node concepts share the same value, except 'objective', 'subjective' and 'alteration', which are more peripheral,

the roles of which might be elucidated with additional network analysis at a mesoscopic level. In general, betweenness centrality is an essential measure, and for researchers in education and cognitive science, since it is considered to be indicative of significant and critically connective concepts that could bridge distant nodes in a belief system, they could ultimately be informative about how certain conceptual entities, i.e., some epistemological beliefs, are formed [50].

Of course, the above analysis did not attempt an exhaustive presentation of all the Gephi capabilities in network analysis. Nevertheless, the choice of a network presentation, likened with the choice of the measures selected, might reveal different aspects of the causal belief system [51], and this indicates that the role of the researcher in the respective choices becomes central, as well. In the present analysis at the microscopic level, the importance of a node in the network is examined through the measures of degree, closeness centrality and betweenness centrality [51]. The centrality measure of the node concepts highlighted the following: "social", "reality", "subjective", "interaction" and "language"; this means that the nodes in question are more influential and determining factors in the network. Noteworthy is the observation that the concepts of "objective" and "subjective" are very close and strongly interrelated in the network, the interpretation of which suggests that in participants' collective mental models, the two ontological assumptions are either indistinguishable or both coexist as kinds of reality.

6. An Outline of the Proposed Methodology

The basic assumption when applying the present methodology is that a belief system can be represented as a complex network of concepts, which reflects the internal conceptual structure of an individual. In this paper, students' epistemological beliefs were the subject under investigation, and it was chosen to illustrate the proposed analytical process. The access to the internal representations, of course, could be achieved by a direct approach, where the participants are asked to draw by themselves those conceptual maps, which are analyzed accordingly (e.g., CmapTools) [16,17]. However, the interest here, is on concept maps originating from content analysis and ensued via computer-aided data mining process. So, in this endeavor text data sources were used, rather than directly formed concept maps, to provide the opportunity for an ampler development of the subjects' reasoning. Moreover, via written essays, the formation of epistemological beliefs does not involve a quick explicit depiction of the assumed relationships, but it is rather a latent path of development, where it is likely to preserve high levels of self-consciousness in externalizing the cognitive system under investigation.

Leximancer software was chosen because it facilitates content analysis in natural language processing and the extraction of conceptual maps from the data, having some advantages, such as, easy-to-use and the option to operate in different languages. The data mining procedures have been for a long time implemented and the algorithms used are greatly improved, so their effectiveness and the validity issues [38,52] have been satisfactorily ensured, and in this endeavor are assumed given.

In addition, for the quantitative part of the methodology, the Gephi software was chosen as a friendly, popular and open-source tool, which includes most of the useful calculation-options for network analysis. Next, we summarize the whole process by describing the main steps of the analysis:

- A. After the textual data were collected the files are prepared according to research questions posited. For example, if one wishes to focus on gender, age, field of study, time period of data collection, etc., then he/she can segment the textual data accordingly in order to perform separate analyses and comparisons. In case, the interest is not focusing on possible individual differences, the accumulated individual files are entered including all subjects' responses. Once the preparation of the files was completed, the process of analysis begins, which includes the following stages.
 - *Document selection*. In this stage, the necessary files or folders to be analyzed are entered.

- Generate concept seeds. This step includes two sub-sets, text processing settings and concept seeds settings. In the first one, it is possible to select settings for the segmentation of the text from which coding will derived, while in the second one, parameters for the identification of concepts, the percentage and the number of concepts can be set.
- *Generate thesaurus*. This stage also includes two sub-processes. In the first, the options of reviewing, removing, or merging the automatically identified concepts is provided. In addition, at this level, concepts can be entered by the user in order to be searched and identified. The second sub-procedure concerns the thesaurus settings that deepen the options previously made on the concept seeds. At this stage, some interventions were made. For instance, concept-words which does not add any specific meaning, such as is, to, concern, example, circumstance, etc. can be removed. Then, merging of concepts in plural–singular, such as science-sciences, subject-subjects, or common words for genders, such as he-she, can also be carried out.
- *Generate concept map*. The fourth stage concerns settings related to the final result, the visualization of the concept map. In the concept coding, concepts can be selected to be included or hidden from the map, and in addition other aspects such as the type of network or the size of the theme, can be selected.
- Qualitative interpretation of the emerged concept map is carried out.
- B. The next phase concerns the preparation and the realization of the quantitative network analysis.
 - *Export* the collective concept map and the corresponding matrices with statistical measures, in csv file.
 - Introduce the matrix file into Gephi for further network analysis.
 - *The network analysis* sought a deeper understanding of the role of each conceptnode in the semantic network. After the matrix has been imported and indicating the number of nodes and edges, the network analysis is specified and focused at the micro-level. Therefore the measures of degree, closeness centrality, and betweenness centrality are implemented, while each measure activates the corresponding feature in the network visualization.
 - *Based* on the quantitative centrality measures the role and the significance of each concept-node in the semantic network is determined.
 - Draw conclusions and develop perspectives.

It is imperative to repeat here that the researcher is not a passive observer, in the data mining procedure, but he/she is an active contributor, who actually could influence the process and the outcomes. Thus, his/her theoretical knowledge and experience are decisive. In addition, the tools, even effective, have also limitations and are amenable to improvement.

7. Discussion

This paper has illustrated a software-aided methodology for content analysis by implementing the Leximancer software package. This is an advantageous approach in data mining starting with plain texts, analyzing them via sophisticated statistical processes and converting them into concept maps [38–40]. This comprises a semantic analysis, since the core concepts and their interconnectivity with other conceptual entities reveal a structure reflecting the mental representations originating from the textual data under investigation.

The specific application presented from education research, probing university students' epistemological beliefs, showed that social, interaction, construction, reality and subjective are the prevailing conceptual elements in the semantic structure. These are expected to be present in a semantic network that potentially adheres to a constructivistic stance. This finding has, of course, certain implications for social science research, at a practical and teaching level, as it was mentioned in a preceding section. However, they will not be discussed, since the scope of this endeavor is the methodological issues of content/semantic analysis. Focusing on this, an auxiliary network analysis at the micro level was included in the methodology by sequentially implementing another specialized program. Gephi performed calculations of quantitative measures such as degree, closeness centrality and betweenness centrality and supported the central role of the above crucial concepts that characterize the networks of meaning under study.

This study, aiming to explore the applicability of software-aided procedures in content and semantic analysis, demonstrated that this methodology in education research deserves special attention for three reasons. First, the advantages of easy-to use, fast and effective tools in processing huge corpora of data certainly comprise the added value of this digital innovation. Second, the semantic analysis can be used within either the traditional qualitative approaches [36,52] or within the quantitative domain by implementation of calculated measures [32–34].

Furthermore, the third reason is that this methodological approach is not a guileless process offered by digital invocations, but it is driven by a theory-supported framework. Besides the advanced technology implementation, this approach posits epistemological issues and could initiate further discussion concerning the methodological framework of social sciences. Since it could be seen as working at the interface of qualitative vs. quantitative methodologies, it adds a problematization concerning a potential unified view. As it was mentioned in a preceding section, the network representation entails the ontology of complex systems, and the conceptual associative networks of meaning are considered as such. The main thesis of the paper is that this *content analysis* actually bridges the traditional dichotomy of qualitative vs. quantitative existing in the methodological landscape in social sciences by fostering the complexity theory view of reality [23–25,27]. That is, the qualitative characteristics of networks of meaning explored here at a macro level *emerge* from quantitative variations in relationships at a lower/micro level of complexity. In other words, the cumulative or quantitative relationships among the network elements determine the macro structure, which is being qualitatively evaluated and interpreted within the current theoretical perspectives.

Author Contributions: The two authors have equally contributed to the present research, sharing the work in each step and the corresponding responsibility. Conceptualization, M.G. and D.S.; methodology, M.G. and D.S.; software, M.G. and D.S.; validation, M.G. and D.S.; formal analysis, M.G. and D.S.; investigation, M.G. and D.S.; resources, M.G. and D.S.; data curation, M.G. and D.S.; writing—original draft preparation, M.G. and D.S.; writing—review and editing, M.G. and D.S.; visualization, M.G. and D.S.; supervision, D.S.; project administration, D.S.; funding acquisition, M.G. and D.S.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was contacted according to guidelines of the Declaration of Helsinki that provides all needed recommendations for carrying out research with adults. The study has been approved by the Ethics and Deontology Committee of the Aristotle University of Thessaloniki (No. 280526/2021).

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank Ioannis Katerelos for the constructive communication on the implementation of Gephi software.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Elements from the concept map for the ontological question including the three central themes and the concepts contained in each, as well as the concepts that co-occur with the central theme.

Main Themes and Frequency of Occurrence	Concepts	Relevance	Coexistence of the 1st Concept with Other Concepts	Estimation of the Conditional Probability
	Reality	100%	social (2092)	100%
	Objective	8%	people (1576)	100%
	Nature	5%	interaction (275)	100%
	Alteration	5%	construction (214)	100%
Poolity (2172)			objective (178)	100%
Reality (2173)			rules (187)	100%
			subjective (119)	100%
			nature (108)	100%
			alteration (98)	100%
			language (67)	100%
	Social	96%	interaction (275)	100%
	People	73%	construction (214)	100%
	Interaction	13%	rules (187)	100%
	Rules	9%	people (1564)	99%
G : 1 (2002)			subjective (118)	99%
Social (2092)			nature (107)	99%
			objective (176)	99%
			alteration (96)	98%
			language (65)	97%
			reality (2092)	96%
Construction (214)	Construction	10%	interaction (68)	25%
	Subjective	5%	subjective (26)	22%
	Language	3%	objective (34)	19%
			language (11)	16%
			people (193)	12%
			social (214)	10%
			reality (214)	10%
			alteration (9)	9%
			nature (8)	7%
			rules (12)	6%

Table A2. Elements from the concept map for the ontological question including the details of the other concepts that are not central themes but are included in them.

Concepts and Frequency of Occurrence	Relevance	Coexistence of Concept with Other Concepts	Estimation of the Conditional Proba-bility
people (1576)	73%	interaction (271)	99%
		construction (193)	90%
		language (59)	88%
		subjective (103)	87%
		rules (159)	85%
		objective (146)	82%
		social (1564)	75%
		reality (1576)	73%
		alteration (67)	68%
		nature (62)	57%

Concepts and Frequency of Occurrence	Relevance	Coexistence of Concept with Other Concepts	Estimation of the Conditional Proba-bility
	13%	Construction (68)	32%
		language (14)	21%
		subjective (23)	19%
		people (271)	17%
interaction (075)		objective (25)	14%
Interaction (275)		social (275)	13%
		reality (275)	13%
		nature (13)	12%
		rules (20)	11%
		alteration (8)	8%
	9%	alteration (13)	13%
		language (7)	10%
		people (159)	10%
		social (187)	9%
		reality (187)	9%
rules (187)		objective (15)	8%
		interaction (20)	7%
		construction (12)	6%
		nature (6)	6%
		subjective (6)	5%

Table A2. Cont.

References

- Burgess, C.; Lund, K. Modelling Parsing Constraints with High-dimensional Context Space. Lang. Cogn. Process 1997, 12, 177–210. [CrossRef]
- Landauer, T.K.; Foltz, P.W.; Laham, D. An introduction to latent semantic analysis An Introduction to Latent Semantic Analysis. Discourse Process. 1998, 25, 259–284. [CrossRef]
- 3. Smith, A.E. Machine mapping of document collections: The Leximancersystem. In *Paper Presented at the Proceedings of the Fifth Australasian Document Computing Symposium*; QLD: Sunshine Coast, Australia, 2000.
- 4. Benzécri, J.-P. Correspondence Analysis Handbook, 1st ed.; Marcel Dekker: New York, NY, USA, 1992.
- Hagiwara, M. Self-organizing concept maps. In Proceedings of the 1995 IEEE International Conference on on Systems, Man and Cybernetics. Intelligent Systems for the 21st Century, Vancouver, BC, Canada, 22–25 October 1995. [CrossRef]
- Mothe, J.; Chrisment, C.; Dkaki, T.; Dousset, B.; Karouach, S. Combining mining and visualization tools to discover the geographic structure of a domain. *Comput. Environ. Urban Syst.* 2006, 30, 460–484. [CrossRef]
- Barrière, C.; Agbago, A. TerminoWeb: A Software Environment for Term Study in Rich Contexts. In Proceedings of the Conference on Terminology, Standardisation and Technology Transfer, Beijing, China, 25–26 August 2006.
- McKenna, B.; Waddell, N. Media-ted political oratory following terrorist events—International political responses to the 2005 London bombing. J. Lang. Polit. 2007, 6, 377–399. [CrossRef]
- Cretchley, J.; Gallois, C.; Chenery, H.; Smith, A. Conversations between carers and people with schizophrenia: A qualitative analysis using leximancer. *Qual. Health Res.* 2010, 20, 1611–1628. [CrossRef]
- 10. Cretchley, J.; Rooney, D.; Gallois, C. Mapping a 40-year history with leximancer: Themes and concepts in the journal of cross-cultural psychology. *J. Cross. Cult. Psychol.* **2010**, *41*, 318–328. [CrossRef]
- 11. Penn-Edwards, S. I know what literacy means: Student teacher accounts. Aust. J. Teach. Educ. 2011, 36, 15–32. [CrossRef]
- 12. Stockwell, P.; Colomb, R.M.; Smith, A.E.; Wiles, J. Use of an automatic content analysis tool: A technique for seeing both local and global scope. *Int. J. Hum. Comput. Stud.* **2009**, *67*, 424–436. [CrossRef]
- 13. Siew, C.S.Q. Using network science to analyze concept maps of psychology undergraduates. *Appl. Cogn. Psychol.* **2019**, *33*, 662–668. [CrossRef]
- Gray, S.A.; Zanre, E.; Gray, S.R.J. Fuzzy Cognitive Maps as Representations of Mental Models and Group Beliefs. In *Fuzzy Cognitive Maps for Applied Sciences and Engineering: From Fundamentals to Extensions and Learning Algorithms*; Papageorgiou, E.I., Ed.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 29–48.
- 15. Novak, J.D.; Cañas, A.J. Theoretical Origins of Concept Maps, How to Construct Them, and Uses in Education. *Reflecting Educ.* **2007**, *3*, 29–42.
- Cañas, A.J.; Hill, G.; Carff, R.; Suri, N.; Lott, J.; Gómez, G.; Eskridge, T.C.; Arroyo, M.; Carvajal, R. CmapTools: A Knowledge Modeling and Sharing Environment. In Proceedings of the First International Conference on Concept Mapping, Pamplona, Spain, 14–17 September 2004.

- 17. Novak, J.D.; Cañas, A.J. The origins of the concept mapping tool and the continuing evolution of the tool. *Inf. Vis.* **2006**, *5*, 175–184. [CrossRef]
- 18. Doyle, J.K.; Ford, D.N. Mental models concepts for system dynamics research. Syst. Dyn. Rev. 1998, 14, 3–29. [CrossRef]
- 19. Carley, K.M. Extracting team mental models through textual analysis. J. Organ. Behav. Manag. 1997, 18, 533–558. [CrossRef]
- 20. Carley, K.; Palmquist, M. Extracting, Representing, and Analyzing Mental Models. Soc. Forces 1992, 70, 601–636. [CrossRef]
- 21. McLinden, D. Concept maps as network data: Analysis of a concept map using the methods of social network analysis. *Eval. Program Plann.* **2013**, *36*, 40–48. [CrossRef] [PubMed]
- Krippendorf, K. Content Analysis: An Introduction to Its Methodology, 2nd ed.; Sage Publications, Inc.: California, CA, USA, 2004; pp. 281–298.
- Stamovlasis, D.; Koopmans, M. Editorial Introduction: Education is a Dynamical System. Nonlinear Dynamics. *Psychol. Life Sci.* 2014, 18, 1–4.
- Koopmans, M.; Stamovlasis, D. Introduction to education as a complex dynamical system. In *Complex Dynamical Systems in Education: Concepts, Methods and Applications*; Koopmans, M., Stamovlasis, D., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 1–7.
- Stamovlasis, D. Catastrophe Theory: Methodology, Epistemology, and Applications in Learning Science. In Complex Dynamical Systems in Education: Concepts, Methods and Applications; Koopmans, M., Stamovlasis, D., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 141–175.
- Stamovlasis, D. Nonlinear Dynamical Interaction Patterns in Collaborative Groups: Discourse Analysis with Orbital Decomposition. In *Complex Dynamical Systems in Education: Concepts, Methods and Applications;* Koopmans, M., Stamovlasis, D., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 273–297.
- Koopmans, M. Mixed Methods in Search of a Problem: Perspectives from Complexity Theory. J. Mix. Methods Res. 2016, 11, 16–18.
 [CrossRef]
- 28. Tsiotas, D.; Polyzos, S. The Complexity in the Study of Spatial Networks: An Epistemological Approach. *Netw. Spat. Econ.* **2018**, 18, 1–32. [CrossRef]
- Bruun, J.; Evans, R. Network Analysis as a Research Methodology in Science Education Research. *Pedagogika* 2018, 68, 201–217. [CrossRef]
- Marion, R.; Schreiber, C. Evaluating Complex Educational Systems with Quadratic Assignment Problem and Exponential Random Graph Model Methods. In *Complex Dynamical Systems in Education: Concepts, Methods and Applications*; Koopmans, M., Stamovlasis, D., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 177–201.
- 31. Katerelos, I.; Koulouris, A. Is Prediction Possible? Chaotic behaviour of Multiple Equilibria Regulation Model in Cellular Automata Topology. *Complexity* **2004**, *10*, 23–36. [CrossRef]
- 32. Stella, M. Mapping the Perception of "Complex Systems" across Educational Levels through Cognitive Network Science. *Int. J. Complex. Educ.* 2020, *1*, 71–89.
- 33. Sun, H.; Yen, P.; Cheong, S.A.; Koh, E. Network science approaches to education research. Int. J. Complex. Educ. 2020, 1, 121–149.
- 34. Koponen, I.T. Complexity and Simplicity of University Students 'Associative Knowledge of History of Science: Knowledge Organisation Poised on the Edge Chaos? *Int. J. Complex Syst.* **2021**, *2*, 3–26.
- 35. Smith, A.E. Automatic extraction of semantic networks from text using leximancer. In Proceedings of the HLT-NAACL Demonstrations, Edmonton, AB, Canada, 27 May–1 June 2003. [CrossRef]
- Harwood, I.A.; Gapp, R.P.; Stewart, H.J. Cross-check for completeness: Exploring a novel use of leximancer in a grounded theory study. *Qual. Rep.* 2015, 20, 1029–1045. [CrossRef]
- 37. Penn-Edwards, S. Computer aided phenomenography: The role of leximancer computer software in phenomenographic investigation. *Qual. Rep.* 2010, *15*, 252–267. [CrossRef]
- Smith, A.E.; Humphreys, M.S. Evaluation of unsupervised semantic mapping of natural language with Leximancer concept mapping. *Behav. Res. Methods* 2006, 38, 262–279. [CrossRef] [PubMed]
- Thomas, D.A. Searching for significance in unstructured data: Text mining with leximancer. *Eur. Educ. Res. J.* 2014, 13, 235–256. [CrossRef]
- 40. Hostager, T.J.; Voiovich, J.; Hughes, R.K. Hearing the Signal in the Noise: A Software-Based Content Analysis of Patterns in Responses by Experts and Students to a New Venture Investment Proposal. *J. Educ. Bus.* **2013**, *88*, 294–300. [CrossRef]
- Leximancer. Leximancer User Guide. 2021. Release 4.5. pp. 1–136. Available online: https://static1.squarespace.com/ static/5e26633cfcf7d67bbd350a7f/t/60682893c386f915f4b05e43/1617438916753/Leximancer+User+Guide+4.5.pdf (accessed on 10 March 2021).
- 42. Miller, K.I. Common ground from the post-positivist perspective: From the "straw person" argument to collaborative coexistence. In *Perspectives on Organizational Communication: Finding Common Ground;* Corman, S.R., Poole, M.S., Eds.; Guilford Press: New York, NY, USA, 2000; pp. 46–67.
- 43. Moore, R. Going critical: The problem of problematizing knowledge in education studies. *Crit. Stud. Educ.* 2007, 48, 25–41. [CrossRef]
- 44. Mack, L. The philosophical underpinnings of educational research. *Polyglossia* 2010, 19, 5–11.
- 45. Scotland, J. Exploring the philosophical underpinnings of research: Relating ontology and epistemology to the methodology and methods of the scientific, interpretive, and critical research paradigms. *Engl. Lang. Teach.* **2012**, *5*, 9–16. [CrossRef]

- 46. Furlong, P.; Marsh, D. A Skin Not a Sweater: Ontology and Epistemology in Political Science. In *Theory and Methods in Political Science. Political Analysis*; Marsh, D., Stoker, G., Eds.; Palgrave: London, UK, 2010; pp. 184–211. [CrossRef]
- Schommer-Aikins, M.; Easter, M. Ways of knowing and epistemological beliefs: Combined effect on academic performance. *Educ. Psychol.* 2006, 26, 411–423. [CrossRef]
- 48. Paulsen, M.B.; Feldman, K.A. Student Motivation and Epistemological Beliefs. New Dir. Teach. Learn. 1999, 17–25. [CrossRef]
- 49. Borgatti, S.; Halgin, D. On Network Theory. SSRN Electron. J. 2013, 1–14. [CrossRef]
- 50. Siew, C.S.Q. Applications of Network Science to Education Research: Quantifying Knowledge and the Development of Expertise through Network Analysis. *Educ. Sci.* 2020, *10*, 101. [CrossRef]
- 51. Siew, C.S.Q.; Wulff, D.U.; Beckage, N.M.; Kenett, Y.N. Cognitive network science: A review of research on cognition through the lens of network representations, processes, and dynamics. *Complexity* **2019**, 2019, 1–24. [CrossRef]
- 52. Lemon, L.L.; Hayes, J. Enhancing Trustworthiness of Qualitative Findings: Using Leximancer for Qualitative Data Analysis Triangulation. *Qual. Rep.* **2020**, *3*, 604–614. [CrossRef]