Lexicons of Key Terms in Scholarly Texts and Their Disciplinary Differences: From Quantum Semantics Construction to Relative-Entropy-Based Comparisons

Koponen, Ismo

Koponen, I.; Södervik, I. Lexicons of Key Terms in Scholarly Texts and Their Disciplinary Differences: From Quantum Semantics Construction to Relative-Entropy-Based Comparisons. Entropy 2022, 24, 1058.

Article

Lexicons of Key Terms in Scholarly Texts and Their Disciplinary Differences: From Quantum Semantics Construction to Relative-Entropy-Based Comparisons

Ismo Koponen [1,*,†] and Ilona Södervik [2,†]

1   Department of Physics, University of Helsinki, 00014 Helsinki, Finland
2   Centre for University Teaching and Learning (HYPE), University of Helsinki, 00014 Helsinki, Finland; ilona.sodervik@helsinki.fi
*   Correspondence: ismo.koponen@helsinki.fi
†   These authors contributed equally to this work.

Abstract: Complex networks are often used to analyze written text and reports by rendering texts in the form of a semantic network, forming a lexicon of words or key terms. Many existing methods to construct lexicons are based on counting word co-occurrences, having the advantage of simplicity and ease of applicability. Here, we use a quantum semantics approach to generalize such methods, allowing us to model the entanglement of terms and words. We show how quantum semantics can be applied to reveal disciplinary differences in the use of key terms by analyzing 12 scholarly texts that represent the different positions of various disciplinary schools (of conceptual change research) on the same topic (conceptual change). In addition, attention is paid to how closely the lexicons corresponding to different positions can be brought into agreement by suitable tuning of the entanglement factors. In comparing the lexicons, we invoke complex network-based analysis based on exponential matrix transformation and use information theoretic relative entropy (Jensen–Shannon divergence) as the operationalization of differences between lexicons. The results suggest that quantum semantics is a viable way to model the disciplinary differences of lexicons and how they can be tuned for a better agreement.

Keywords: quantum semantics; semantic networks; lexicons; text analysis

1. Introduction

Research fields always contain disciplinary groups where key scientific terms differ, and the same terms may be used differently in discussing and framing the key problems within the field. Such disciplinary fragmentation is particularly characteristic in the human and behavioral sciences [1,2]. Since scholarly texts are a major medium for disseminating the ideas and views of disciplinary groups, examination of how key terms are used in them provides important information on disciplinary differences. In this study, we suggest a simple method suitable for constructing, analyzing, and comparing networks of key terms (lexicons) in 12 scholarly texts related to learning sciences, to examine the differences and to find out how the lexicons could be tuned for better overlap.

The disciplinary structure of science and the formation of disciplinary schools, along with their characteristic ways of using scientific terms and forming various research programs, have long been topics of interest [3–5]. The use of networks as representations of the structure of science and its conceptual systems [4,5] was also recognized quite early, decades before the science of science [4] matured into the new research field of the cartography of sciences [6–8]. Contemporary science of science [9,10] has provided a variety of means to study the structure of sciences, their disciplinary and conceptual structures [6–8], and means for characterizing disciplinary differences [9–14]. Many such studies utilize methods based on text analysis and linguistic approaches developed within network science [15].

Focus on linguistic structures is also supported by studies within the philosophy and history of science. The language and linguistic structures involved in forming disciplinary groups, and their identity and ways of communicating scientific views and ideas, were central themes in Kuhn's conception of science, which he developed after his views about scientific revolutions and paradigms [16]. In these later studies, Kuhn focused on the relational structures of scientific terms and concepts and how a networked, interlinked structure of concepts supports the formation of their meaning, referring to such networks as lexicons or lexical networks [16–18]. According to Kuhn, a lexicon is a network of terms, and terms are related through their connections in context. Here we adopt this view and take lexicon to correspond to the relatedness of words through their co-occurrence in texts, or rather, as will be seen later, how the co-occurrences are entangled in text. The basic assumption here is that disciplinary differences emerge from users of the lexicons having the option to use different interrelationships within the adopted lexicon. Such lexicons thus become basic structures characterizing the disciplinary groups and their identities.

Taking lexicons of scientific terms central for identifying disciplinary groups lends itself easily to the idea of using scholarly texts as sources to explore the similarities and differences in the use of terms between disciplinary groups within the same research field [15]. The semantic structure of texts and the semantic fields attached to their key terms are often approached using some version of co-word mapping or bag-of-words methods (see, e.g., reference [15] and references therein). Co-word mapping methods have a long history going back to seminal studies in the 80s and 90s [19–23], then still hampered by limited computing resources [15]. In the last two decades, co-word mapping methods have given room to the now widely applied topic modeling of texts, which has its origins in natural language processing and attempts to find topics on basis of analyzing bags of words from the text [15,24–26]. Consequently, word co-occurrence analysis and its different variants still find new and innovative applications in text and text-structure analysis [27–30].

Recently, text analysis methods that are more sensitive than co-word mapping to text structure have been introduced, with the aim of revealing syntactic and semantic structures that run deeper than co-word and semantic field analyses. The co-word mapping methods capture syntactic structures only up to word adjacency, but they can be generalized to take into account more complicated syntactic structures composed of subject–verb–object triads and networks constituted by such triads [31]. Other methods going beyond co-word analysis are looking in detail for example: concept diversity (cognitive content) through lexical diversity of text [14]; relation of key concepts and the architecture of the text structure [32,33]; searching semantic frames through finding communities of verbs and their arguments [34]; creating semantic networks by using concept-centered sub-networks [35]; and finding the larger-scale semantic structure of texts [36,37]. In all these cases, complex networks methods are used in construction of the semantic, lexical, or concept networks, and different network measures (closeness centrality, betweenness centrality, or some specially engineered measures for network topology) are utilized in analysis. Moreover, different types of computer software exist for different kinds of analysis methods—for example, for co-word mapping and topical analysis (for a reviews, see reference [15,37]) and analysis of semantic frames (for a review, see reference [34]). Here, however, the focus is not on the comparison of different methods or on providing a review of available methods and software, and therefore, more detailed discussion of such approaches and methods is beyond the scope of this study.

In this study, we propose a generalization of word co-occurrence counting to construct a network of key terms, a kind of lexical network, where links connecting key terms represent the probability of (generalized) co-occurrence. Following Kuhn, we refer such lexical networks as lexicons. The generalization we are proposing is meant to address the possibility of subjective differences in reading the meaning of a given sentence in a text. One potential way to overcome this shortcoming is a recently introduced method called

quantum semantics, which provides a means to estimate the effects of subjective bias and entangled meanings [38–41].

Quantum semantics can be seen as a generalization of word co-occurrence counting methods or bag-of-words methods. It preserves the simplicity of co-word counting while allowing the inclusion of a factor to estimate the effect of subjective bias in reading the meaning of words in a bag of words [38–41]. From the perspective of quantum semantics, semantic fields are not fixed or unyielding structures that define relationships between words as extracted from texts, but rather, tentative templates for how producers of texts can be expected to compose sentences, and how readers of text can be expected to weigh the importance of co-occurrence of words in a sentence, i.e., read meaning into a sentence. In this study, we use quantum semantics in constructing the semantic networks (i.e., lexicons) that allow us to explore the extent of disciplinary differences in lexicons.

Seminal ideas of quantum semantics appeared already two decades ago and were applied to word association tasks and categorization tasks (see, e.g., reference [41] for a review), based on the insight that word meanings are entangled in the sense that the context where they appear with other words affects the ability to factorize meanings of words to elementary, independent elements (i.e., kinds of pure states). To describe such entangled states of word pairs, formalism from quantum logic, essentially in form of qubits, was adopted [41]. The recent idea to quantify the entanglement by using a quantum information theoretic measure called concurrence [38–40] provides substantial simplification to adopt quantum semantics as the basis to construct lexical networks. It should be noted that adopting quantum semantics to generalize word co-occurrence counting does not imply endorsing a view that words are quantum entities, only that generalized logic and probability behind the quantum theory are adopted. Here, we use such an approach to analyze scholarly texts to construct lexicons (lexical networks) of their key terms.

Quantum semantics makes it possible to study effective subjective bias between word pairs, and with that knowledge, to build a lexicon in the form of a (semantic) network. To explore the structure of lexicons, we invoke methods developed for analysis of complex networks. A network description is here based on a (weighted) adjacency matrix obtained through quantum semantics. The analysis is then carried out using an exponential matrix transformation that allows us to define a density matrix [42–46] for the lexicon (i.e., the probability of connectivity between links). The density matrix, on the other hand, when suitably scaled, can be interpreted as a correlation matrix between nodes in the network [47], i.e., now as correlations between co-occurrence of words when each pair of words is assumed to be entangled. Note that concurrence as a measure of entanglement between a pair of words or terms is the basis of obtaining the correlation matrix, but correlation matrix, as it is based on density matrix, takes all connections contained in the network into account.

The density matrix also allows a straightforward quantitative method for comparison of the lexicons, to be based directly on entropy differences between different lexicons. Here, the symmetric form of a relative entropy (here von Neumann entropy, because we are working with density matrix) in the form of Jensen–Shannon divergence [48–50] is adopted as the basis of comparison. Owing to its symmetric form and computational stability, Jensen–Shannon divergence is a convenient choice in comparing semantic networks and word distributions. Moreover, the square root of Jensen–Shannon divergence can be interpreted as a kind of (information) distance metric between networks [51–53].

To demonstrate the viability of this method based on quantum semantic networks in characterizing disciplinary differences of lexicons, we have chosen 12 scholarly texts concerning models of conceptual change (see reference [54] for a review). The texts deal with conceptual change models from different and even contradictory viewpoints. These texts are particularly suitable because the discrepancy of views they provide, to a shared research theme, is well known and discussed in the research literature. Therefore, it should be interesting for the reader to see if the method of analysis proposed here will

reproduce the expected differences in level for lexicons, and in addition, whether or not the possibility to tune for better agreement of lexicons affects their similarity in essential and substantial ways.

It is shown that the method proposed here, although quite simple, indeed brings out clear differences between the texts, but also provides an understanding of how, through subjectively biasing a better overlap of their lexicons, mutual agreement could be possible. We believe that the methods introduced here may support the common traditional interpretative analyses, providing simple but sufficiently reliable quantitative measures to detect disciplinary differences in scholarly texts, and also providing a means to estimate a range of variation in differences.

2. Materials and Methods

The method to construct and compare lexicons consists of three methodologically different but closely interrelated steps:

1. Quantum semantics is used to construct the lexicons corresponding a given text.
2. The relevant connectivities of lexicons are found through network analysis.
3. The comparison of lexicons is performed based on a quantum information theoretic version of Jensen–Shannon divergence.

We have limited discussions of methods to the minimum of what is needed to explain the steps in the analysis. The details of the derivation behind the results utilized here are well documented in the original references. The text corpora of scholarly texts used to test the method are 12 texts of conceptual change research. The texts were chosen to correspond to positions known to be seminal (three texts) texts displaying two opposing views (three texts for each case), and texts known to mediate between the opposing views (three texts).

2.1. Text Samples

The text samples chosen were 12 scholarly texts about models of conceptual change [55–66], published between the years 1982 and 2017. Short descriptions of the 12 texts selected for analysis are listed in Table 1 and identified with the acronyms provided. The selection was made on the basis of the widely held opinion that the research field of conceptual change has seen a long-standing dispute about the nature of conceptual change, in which some parties have tried to mediate between different views, but some other disciplinary schools have not found or even striven for a consensus. However, the basic terms and terminology and conceptions of the basic topic of interest are agreed on. In many ways, such a situation offers an interesting test-bed for a quantum semantics approach.

Table 1. Scholarly text about conceptual change used as a sample. Acronyms (Acr.) specify the text provided as a reference, which is briefly described in last column.

Acr.	Ref.	Description of Position
Ca	[55]	Developmental psychology and cognition; theories of concepts; cognitive science
P1	[56]	Piaget's theory of learning; Kuhn's theory of science; scientific knowledge
P2	[57]	Updated and augmented version of P1
V1	[58]	Framework theory; Knowledge-as-theory view; synthetic conceptual models
V2	[59]	Updated and extended version on V1; Contrasts to Knowledge-as-elements view
V3	[60]	Review of views in V1 and V2; Contrasts to Knowledge-as-elements view
D1	[61]	P-prims and knowledge-as-elements view; Contrasts to knowledge-as-theory view
D2	[62]	Augments D1 by Coordination classes; Contrasts to knowledge-as-theory view
D3	[63]	Review of views D1 and D2; Contrasts to knowledge-as-theory view
OC	[64]	Attempts synthesis between knowledge-as-elements and -theory views
A1	[65]	Discusses views Ca, P1-2, V1-2, D1-2 and attempts to find common viewpoint
A2	[66]	Review, augments A1, with specific emphasis on views as in Ca

A well-known and widely discussed dispute is related to the question of whether conceptual knowledge should be seen as forming out of isolated and fragmented pieces (knowledge-in-pieces view according to some authors) or as broader and in some sense coherent theory-like structures (framework theories, according to some authors). In such disputes, referred to in what follows as "knowledge-as-elements" or "knowledge-as-theory" views (for reviews, see references [64,65]), the discussions are about how to frame the meanings of key concepts to be used in conceptual change. In this dispute, advocates of both views have held their positions and continued to discuss and criticize the opposing positions.

Of the twelve texts, texts V1–V3 belong to the camp of "knowledge-as-theory", and texts D1–D3 are in the camp "knowledge-as-elements". The texts OC, A1, and A2 are reviews, in which consolidation of different views is attempted. In particular, the texts A1 and A2 contain pleas for increased attempts to unify the terminology and strive for better mutual understanding in order to advance the research field. The texts Ca, P1, and P2 are seminal works that are often cited [54] but perhaps not so often directly used in developing researchers' own positions. In this study, we do not discuss in greater detail the content of the texts or the disputes. Instead, we are interested in exploring whether the known opposing positions and attempts to mediate between positions are visible at the level of terminology and how terms co-occur in texts, i.e., at the basic level of semantic fields.

2.2. Generalizing Word Co-Occurrence: Concurrence

The first step in transforming the written texts into semantic networks, in the form of lexicons of terms, consist of splitting the sentences into clauses, in order of their appearance, and after that, recognizing the nouns, and within them, term-like words. The texts were first simplified by removing stop words and miscellaneous symbols. Next, nouns, verbs, adjectives, and attributes were picked out from sentences, preserving the sentence structure. Finally, resulting expressions were stemmed. In that, Mathematica 13 employing Porter's algorithm for stemming [67] was used as software. While lacking advanced features of more advanced natural language processing (NLP), Mathematica produced a quite acceptable outcome and was selected here because it is supposedly easy enough to adopt, in comparison to more advanced NLP methods. This choice is a trade-off between accuracy and simplicity of analysis, because it is not likely that too complex and demanding methods will be adopted by researchers not working in field of NLP but who need to do similar kind of analysis in fields of education and behavioral sciences. After stemming, term-like words that appeared only once were discarded, and from the remaining set, about 70 were chosen for closer attention. Consequently, the selection process is quite simple and misses many finer points, but contains basic and robust elements commonly identified in analyzing speech and writing.

The second step consisted of finding the co-occurrence statistics for the 70 selected terms. To do so, we utilized quantum semantics approach [38–40], which generalizes the traditional word co-occurrence measures so that subjective factors can be taken into account at some level of idealization. In that generalization, the notions of entanglement of words and concurrence as a measure of the entanglement play key roles. Details behind the quantum semantic approach, as they are relevant here, are provided in Appendix A. The key result needed here is the concurrence Q, which quantifies the degree of entanglement of terms as they appear in a text. Symbols used in developing the methods are summarized in Table 2.

To obtain concurrence Q (degree of entanglement) between words A and B in a text, we count four different frequencies of co-occurrence: the frequency n_{11}, that A and B both occur in a given block of clauses at least once; n_{00}, that neither A or B occurs; n_{10} that A occurs at least once but B does not occur; and n_{01}, that A does not occur but B occurs at least once. Such a situation can be described in the form of a classical 2×2 contingency table of association of two variables A and B. In quantum semantics, the association is described in terms of qubits, allowing for non-classical dependence due to entanglement

(due to inherent inseparability and impossibility of factorization of co-occurrence). With these frequencies, the concurrence Q as a measure of entanglement is given as [38,39] (see also reference [68])

$$Q = 2\sqrt{n_{11}n_{00} + n_{10}n_{01}}\sqrt{1 - 2\Theta\frac{\sqrt{n_{11}n_{00}n_{10}n_{01}}}{(n_{11}n_{00} + n_{10}n_{01})}}, \quad 0 \le Q \le 1, \tag{1}$$

where $-1 \le \Theta \le 1$ is the phase factor (see Appendix A for details and derivation) taken here as a free parameter to account for subjective bias [38–40]. For further reference, we define the prefactor $Q_0 = 2\sqrt{n_{11}n_{00} + n_{10}n_{01}}$ that corresponds to concurrence obtained with $\Theta = 0$. The last factor in Equation (1) given by $R = 2\sqrt{n_{11}n_{00}n_{10}n_{01}}/(n_{11}n_{00} + n_{10}n_{01})$ is the ratio of the geometric mean of frequencies $n_{11}n_{00}$ and $n_{10}n_{01}$ to their arithmetic mean. With these definitions, concurrence takes a form $Q = Q_0\sqrt{1 - \Theta R}$. The entanglement is now possible only in cases $R \ne 0$.

Table 2. Summary of symbols and abbreviations used recurrently in the text and figures.

Symbol/Abbreviation		Symbol/Abbreviation		Symbol/Abbreviation	
Q	Concurrence	$\mathbf{Q}$	Concurrence matrix	β	Time-like parameter
Q_0	Prefactor of Q	$[\mathbf{Q}]_{ij}$	Element ij of $\mathbf{Q}$	$\mathbf{\Gamma}$	Correlation matrix
R	Ratio of averages in Q	$\mathbf{C}$	Contingency matrix	Γ_i	Correlation centrality
Θ	Phase factor in Q	$[\mathbf{C}]_{ij}$	Element ij of $\mathbf{C}$	$\bar{\Gamma}$	Vector of elements Γ_i
n_{xy}	Co-occurrence frequency	$\mathcal{W}$	Normalized $\mathbf{W}$	H	von Neumann entropy
$\mathbf{W}$	Weighted adjac. matrix	$[\mathcal{W}]_{ij}$	Element ij of $\mathcal{W}$	J	J-S divergence (JSD)
$[\mathbf{W}]_{ij}$	Element ij of $\mathbf{W}$	ρ	Density matrix	S_{JSD}	JSD-similarity
d_i	Weighted node degree	$[\rho]_{ij}$	Element ij of ρ	S_{COS}	cos-similarity

For purposes of comparison, we note an analogy of concurrence to different factors measuring association between dichotomous variables in [69–71] (see also [38]). In classical methods, the deviation of associations corresponding to randomly shuffled and punctuated (random mixing) sentences leads to equivalence $n_{11}n_{00} = n_{10}n_{01}$ (representing the so-called odds ratio of value 1). Consequently, classical measures of contingency are often taken as either proportional to a factor $n_{10}n_{00} - n_{00}n_{11}$ (e.g., mean square contingency or Yule's phi coefficient) or to $\sqrt{n_{01}n_{10}} - \sqrt{n_{00}n_{11}}$ (Yule's Y coefficient). However, different other ways to measure the deviations from odds ratio are also used (see, e.g., [70,71] for a summary). There is no unambiguous way to define a contingency measure, but in all of them, deviations from an odd ratio of 1 (i.e., no correlations) is central. Consequently, in analogy with Equation (1), where phase factors c and frequencies n are related as $c = \sqrt{n}$, we define a modified contingency as

$$C = 2|\sqrt{n_{01}n_{10}} - \sqrt{n_{00}n_{11}}|, \quad 0 \le C \le 1. \tag{2}$$

The modified contingency C as defined in Equation (2) fulfills now the normalization condition $\sum n_{ij} = 1$ and corresponds to the minimum value of concurrence Q obtained with $\Theta = 1$. Therefore, it can be then compared with Q to illustrate the difference in quantum semantics-based quantification of co-occurrence to a minimal value of concurrence, which is closely related to contingency coefficients measuring traditional analysis of dichotomous variables. However, a more detailed quantitative match is not possible between concurrence Q and different classical contingency measures.

2.3. Constructing Lexicons from Concurrences

We are interested in the concurrence Q of all pairs of terms i and j in a set of 70 selected terms of interest. Each pair is then characterized $Q_0(i, j) = Q_0(j, i)$, along with factors $R(i, j) = R(j, i)$. These factors are obtained for all pairs of terms in all texts. In what

follows, we are interested in concurrence $Q(i, j)$, tuned so that in two texts T and T′ the corresponding concurrencies $Q(i, j)$ and $Q'(i, j)$, respectively, become as close to each other as possible; i.e., we choose factors $\Theta(i, j)$ and $\Theta'(i, j)$ so that difference is minimized. In what follows, Q always refers to such optimized concurrences as a best scenario for matching lexicons. The contingency C, on the other hand, corresponding to the minimal concurrence, provides the worst scenario as a benchmark. The pairwise values $Q(i, j)$ and $C(i, j)$ are used to form weighted adjacency matrices Q and C with elements $[\mathbf{Q}]_{ij} = Q(i, j)$ and $[\mathbf{C}]_{ij} = C(i, j)$, respectively, describing the connectivity of terms in the lexicon.

2.4. Characterizing Lexicons: Density Matrix

To explore the structure of lexicons, we introduce the weighted adjacency matrix **W** with elements $[\mathbf{W}]_{ij} = W_{ij}$, where W_{ij} is taken to be either $Q(i, j)$ or $C(i, j)$ for nodes i and j and $W_{ij} = 0$ for unconnected nodes. In what follows, symmetric adjacency matrices are assumed. For a weighted adjacency matrix, it is recommended to use normalization [42,43,49,72] to obtain a normalized weighted adjacency matrix $\mathcal{W}$, which for connected nodes has elements $[\mathcal{W}]_{ij} = W_{ij} / \sqrt{d_i d_j}$, where $d_i = \sum_{j \neq i} W_{ij}$ is the weighted degree (strength) of node, while for unconnected nodes $[\mathcal{W}]_{ij} = 0$.

To assign a probability density to links in the network, we then introduce density matrix $\boldsymbol{\rho}$ characterizing the network [42–46]

$$\boldsymbol{\rho} = Z^{-1} \exp[\beta \mathcal{W}],\tag{3}$$

where $Z = \mathrm{Tr} \exp[\beta \mathcal{W}]$ is a normalization factor. Due to normalization, $\mathrm{Tr}\boldsymbol{\rho} = 1$, the matrix, has an analogous role to probability density; due to its constructions based on a real and symmetric matrix $\mathcal{W}$ is Hermitian; and finally, due to exponential matrix transformation, it is positive semidefinite matrix; therefore, it can be taken as a density matrix. Here, the weight factor β defines how links are weighted, with low values corresponding to situations where the network disintegrates into an unconnected set of nodes, and high values of β emphasize the role of strong links. The density matrix $\boldsymbol{\rho}$ can be also interpreted as path or walk counting, providing information on all paths or walks that connect the nodes in a network [42–46]. The counting of walks is closely related to the question of how nodes can be reached in given networks, through connecting links. Therefore, we can use density matrix in Equation (3) to establish a connection between a correlation matrix, correlating "positions" of nodes in the network, "position" meaning how through different paths nodes can be reached. In what follows, a short summary is provided, and details not essential for the rest of the study are relocated to Appendix B.

The density matrix $\boldsymbol{\rho}$ provides a way to formulate a correlation matrix describing how "positions" of nodes in the network are correlated [47]. Following Estrada's derivation (see Appendix B and reference [47]), one can show that it is possible to define a correlation matrix $\boldsymbol{\Gamma}$ in the form

$$\boldsymbol{\Gamma} = \gamma^{-\frac{1}{2}} \boldsymbol{\rho} \, \gamma^{-\frac{1}{2}},\tag{4}$$

where $\gamma = \mathrm{Diag}[\boldsymbol{\Gamma}]$ is a diagonal matrix. The elements of $[\boldsymbol{\Gamma}]_{ij}$ of the correlation matrix can be shown to be directly related to the covariance of values of nodes i and j in the network (see Appendix B and reference [47]), providing a way to find the key nodes on the basis of correlations. Towards this end, we define a correlation centrality as an off-diagonal sum:

$$\Gamma_i = \sum_{j \neq i} [\boldsymbol{\Gamma}]_{ij},\tag{5}$$

which closely resembles communicability centrality [42,43]. The correlation centrality is used here as a basis to define key terms and their rankings.

2.5. Comparing Lexicons: Divergence and Similarity

Next, we turn to the task of comparing lexicons described by different density matrices ρ and σ. An obvious approach is to use information theoretic relative entropy as a measure of difference (see, e.g., [48–53]). Relative entropy quantifies the amount of information that is needed to make inferences about a given network using information contained in the network as a basis of comparison. Here, we use the Jensen–Shannon divergence (relative entropy), which is widely used in the comparison of semantic networks [51–53]. Most often, in applications to semantic meaning of words, Jensen–Shannon divergence is applied in context of ordinary probability densities [51–53], but can be generalized for density matrices, as far as different entropy functions can be generalized for density matrices (which is usually accepted as in classical case of generalizing Gibbs–Shannon entropy as von Neumann entropy in quantum mechanics) [73–76].

Jensen–Shannon divergence (JSD) is based on von Neumann–Shannon information theoretic entropy $H(\rho) = -\text{Tr}\,\rho\log\rho$ and is defined as a symmetric relative entropy between density matrices σ and ρ in the form [73–76] (see also [48–50])

$$J(\rho||\sigma) = H(\frac{\rho + \sigma}{2}) - \frac{1}{2}(\,H(\rho) + H(\sigma)\,)\,, \;\; 0 \leq J \leq 1\,. \tag{6}$$

In characterizing systems that can be represented in terms of density matrices, Jensen–Shannon divergence has several convenient properties: It is positive, zero if and only if $\rho = \sigma$, symmetric, and always well-defined. In addition, the square root $\sqrt{J(\rho||\sigma)}$ of the Jensen–Shannon divergence can be also interpreted as an information-based metrics for the distance between σ and ρ [75,76]. Nevertheless, in practice, computation of divergence requires some caution, because it involves computing a matrix logarithm. To overcome issues related to the stability of computation, we have used so-called hypoentropy [77]: $H_\lambda(\rho) \propto -[\text{Tr}\,(\mathbf{I} + \lambda\rho)\log(\mathbf{I} + \lambda\rho)]/\lambda$, which has the limiting values $H_\lambda \to H$, when $\lambda \to \infty$, thus providing von Neumann entropy [77].

The similarity of a network described by ρ to a network described by σ can now be defined through Jensen–Shannon divergence (JSD) as

$$S_{\text{JSD}}(\rho||\sigma) = 1 - \sqrt{J(\rho||\sigma)}\,, \;\; 0 \leq S_{\text{JSD}} \leq 1\,, \tag{7}$$

where the square root of JSD is used because it can be interpreted as a metric, as noted before. This kind of a similarity measure is an example of information-theoretic similarity measures (compare, e.g., [78].

For comparisons, we use a different type of similarity, based on correlation centrality values of nodes. By forming for lexicons L and L$'$ centrality vectors $\bar{\Gamma}$ and $\bar{\Gamma}'$ consisting of centralities Γ_i and Γ'_i of nodes, respectively, we can define the so-called cosine similarity S_{COS} as a dot-product [79,80]

$$S_{\text{COS}} = (\,\bar{\Gamma} \cdot \bar{\Gamma}'\,) \,/\, (\,|\bar{\Gamma}||\bar{\Gamma}'|\,)\,, \;\; 0 \leq S_{\text{COS}} \leq 1\,. \tag{8}$$

The cosine similarity S_{COS} amounts to comparing the importance of key terms in lexicons. When it attains a value 1, networks are completely similar, as characterized by vectors of values of correlation centralities of nodes, whereas for completely different networks, cosine similarity has a value 0. As the correlation centralities of nodes are obtained as off-diagonal sums of the correlation matrix, related to density matrix, we can expect that similarity S_{JSD} based on Jensen–Shannon divergence and cosine similarity S_{COS} provide results that are in agreement, although they may differ slightly; they provide complementary information on the origin of the similarity.

3. Results

The method to construct, analyze, and compare lexicons was demonstrated by analyzing 12 scholarly texts related to conceptual change (see Table 1). First, we discuss results on the basic features and statistics of lexicons. Second, results in identifying key nodes are

introduced, and third, on the basis of these results, the similarity of lexicons is discussed. Fourth and finally, key terms responsible for the similarity of lexicons are presented.

3.1. Word Frequency Statistics

The text analysis and stemming was carried out by using Mathematica 13, as explained in previous section. The texts were of different lengths, and they contained different numbers of words and sentences. The numbers of sentences and words after stemming in each text are provided in Table 3. Although the absolute numbers of sentences and words varied quite a lot, the ratio of words to sentences was in range from 8 to 16; the mean value and standard deviation were 12 ± 2. In what follows, only relative frequencies of co-occurrence are of interest; absolute numbers are of no further relevance.

Table 3. Numbers of sentences (N_S) and words (N_W) and their ratio N_W/N_S in texts listed in Table 1.

	Ca	P1	P2	V1	V2	V3	D1	D2	D3	OC	A1	A2
N_S	339	160	453	240	245	86	626	392	105	260	410	287
N_W	3801	1617	3793	3299	3449	1205	5692	3720	1075	3092	6185	3651
N_W/N_S	11.2	10.1	8.4	13.7	14.1	14.0	9.1	9.5	10.2	11.9	15.1	12.7

Of the list of stemmed expression, the most relevant 150 words and terms were selected for further scrutiny. Selection was done on basis what appeared to be of relevance for the topic (conceptual change, learning, and instruction) in question. For example, "earth" and "hollow", though occurring in many documents, were dropped, since they are related to one specific example often discussed, but not of interest to discussing the general nature of conceptual change. From the resulting list, we then selected 90 terms having the highest frequencies of appearance for closer scrutiny, to be analyzed through their co-occurrence statistics.

Eventually, on basis of co-occurrence, only about 50 out of the 90 terms appeared to be are of interest. These terms are also listed in Figures 7 and 8, where we can see that most of them, but not all, are obviously relevant for the topic of conceptual change. The frequency distribution of the 50 most frequently occurring terms in all 12 texts is shown in Figure 1 (panel a, left). The distributions are shown as functions of rank of occurrence; the most common term has the highest rank of 1, and the larger the rank number, the lower the frequency. The term "concept" is omitted, because its appearance is dominant, being about two times more frequent than the next most abundant term, "student." The distribution of terms, being flat, shows how no specific term appears preferentially in the 12 texts chosen as examples. This is also seen in individual distributions (Figure 1b, right, in the same order of rankings as in the large figure a at left) shown separately and denoted by acronyms (see Table 1). As can be seen, a few terms stand out in each individual text as the most abundant, but they are different ones in different texts. This reveals the diversity of texts even just on the basis of word frequency distributions.

3.2. Lexicons: Concurrence and Construction

To get a more in-depth picture, we next turn to the (modified) contingency C and concurrence Q, and construct the lexicons in a form of a network. The concurrence Q is of special interest, because optimizing it allowed us to tune the lexicons to overlap better, and thus led to more similar semantic fields. As was explained previously (Section 2.3), in tuning lexicons for the best overlap, we found for each link a phase factor which brings the concurrences as close to the their common average value as possible (constrained only by a factor R). For each pair of lexicons, and for each pair of links in each lexicon, these phase factors are different.
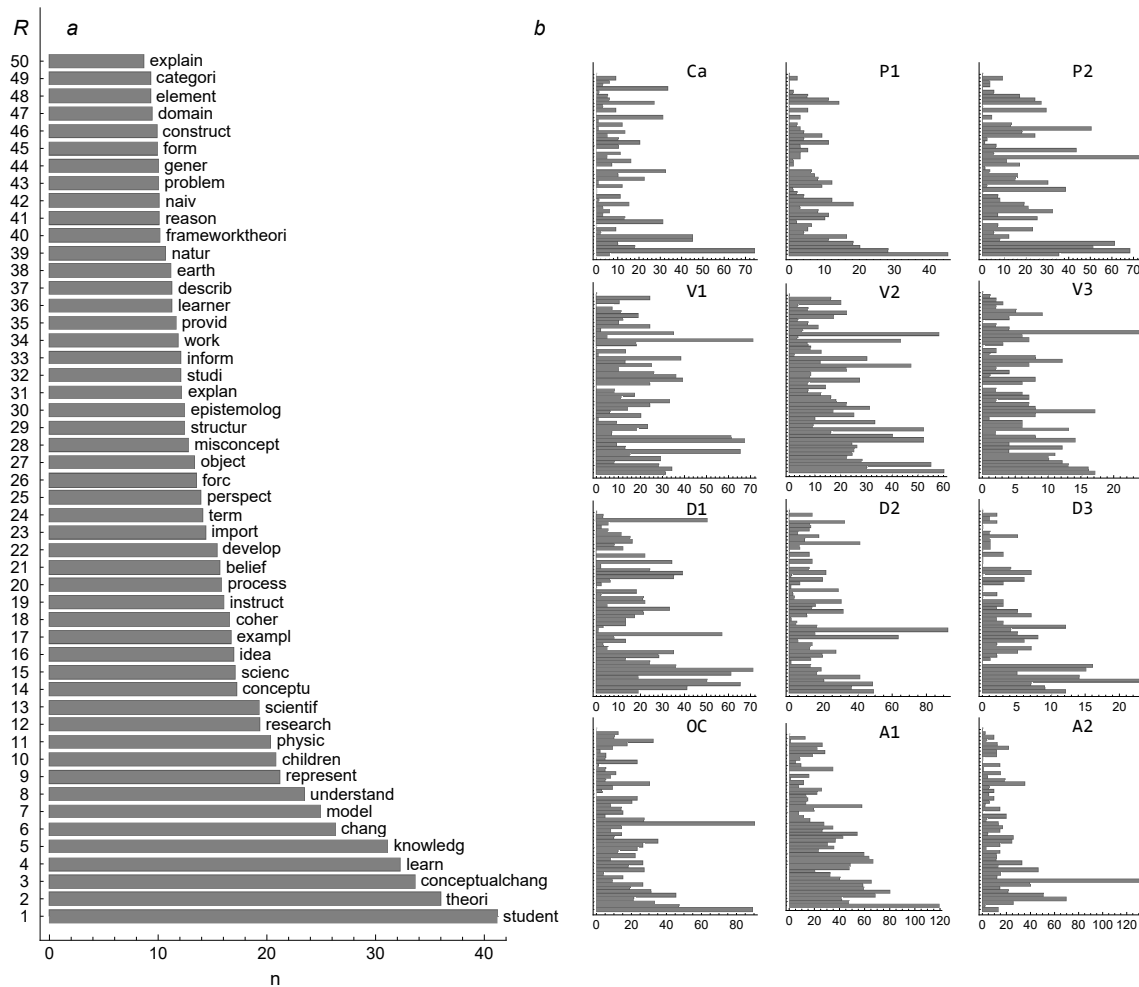
Figure 1. The distributions of occurrence (number) frequency *n* of terms. The rank *R* according to number frequency *n* is shown in panel (**a**) on the left for the 50 most frequently appearing key terms in all 12 texts. The distributions of frequencies in individual texts (denoted by acronyms referring to Table 1) are shown on the right in panel (**b**), in the same order as in (**a**).

Figure 2 shows three examples of how optimized concurrences *Q* and corresponding contingencies *C* are related for lexicons A and B. The examples in Figure 2a–c (in upper row) show that contingencies *C* (gray symbols) for connections between word/term pairs in different lexicons A and B are not related to a significant degree. The situation changes when one turns to optimized concurrence *Q* (black symbols), when for many (but not for all) term pairs it is possible to find a phase factor Θ that brings the concurrences of term pairs in lexicons A and B close to each other; in many cases, values of *Q* in lexicons A and B are aligned, forming a straight line, as seen in Figure 2a–c. This demonstrates that optimization significantly improves the overlap of lexicons, and consequently, the necessary conditions for the formation of shared meanings. For example, in cases (Figure 2a) of A = A1 tuned to B = V2 and D1 (some of the best matching cases), we can see a significant increase in how concurrences are aligned. In the case (Figure 2c) of A = V3 tuned to B = D3 (one of the worst-matching cases), the tuning does not lead to equally high optimal values of *Q*, but still much higher values than obtained for contingency *C*, indicating a significant increase in the overlap of lexicons even in that case. The results in Figure 2 demonstrate the effect of optimization and differences between optimized concurrences and contingencies of term pairs. We will later (Figures 7 and 8, having first discussed the similarity of lexicons) return to the question of which terms are those ones yielding optimization, and at the same time, having significant roles in increasing the similarity of lexicons.

A complementary picture of the effects of tuning is provided in Figure 2d–f, where the same cases as in Figure 2a–c are shown differently, plotting the values of concurrence Q against the values of contingency C in each individual lexicon. In the results in Figure 2d–f, we can observe that while some terms have same values of Q and C, most of the terms have widely different values, and in general, values of Q are higher than those of C (the cases A = V3 and B = D3 being particularly clear). This is, of course, as expected, provided that Q is tunable. In summary, from the results in Figure 2, we can conclude that many terms yield optimization (points above the straight line in Figure 2a–c), and these highly tunable term pairs play an important role in improving the overlap of lexicons and providing shared meanings.
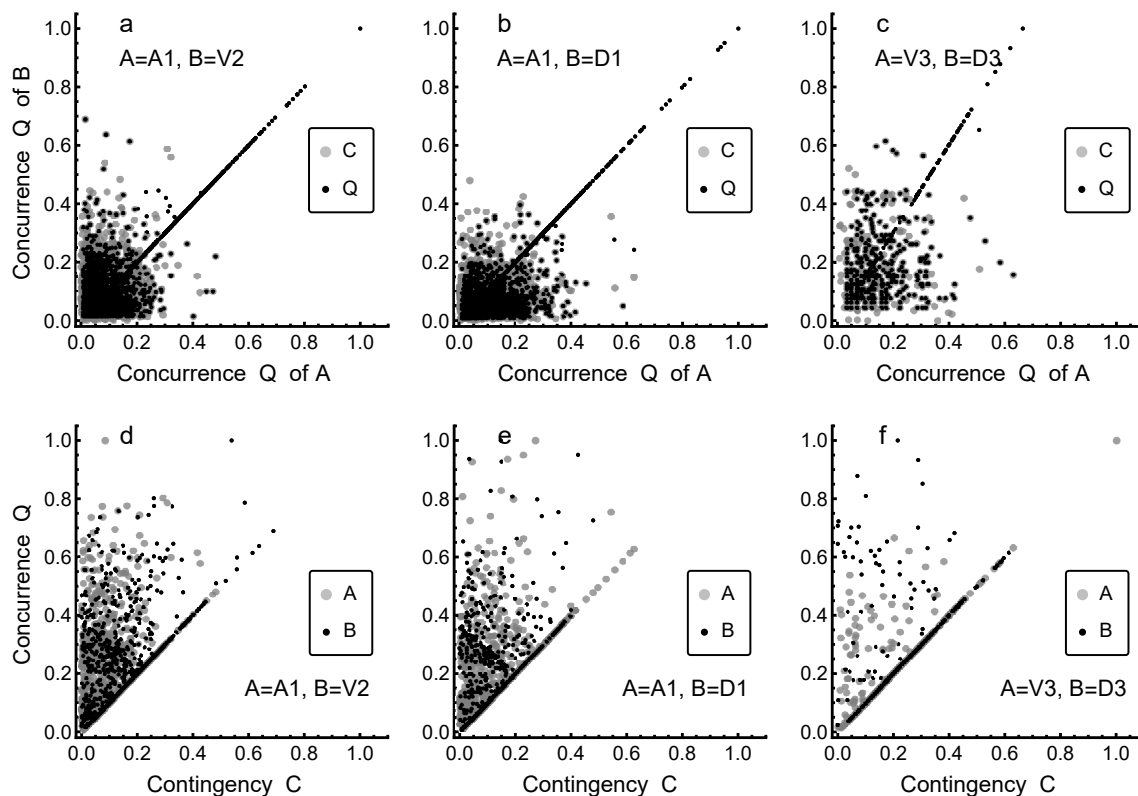

Figure 2. Optimized concurrences' (Q) corresponding contingencies C compared. (**a**–**c**) (upper row) display three examples in the form of a scatter plot showing the optimized concurrences Q (black symbols) and corresponding contingencies C (gray symbols, in the same scale, although the axis label is only for Q) for lexicons A and B. The three cases are indicated by acronyms referring to texts (see Table 1). (**d**–**f**) (lower row) display the same cases as (**a**–**c**) but plot the values of concurrence Q against values of contingency C for each lexicon.

To provide an idea of the diversity of lexicons, we show in Figure 3 all 12 cases as projections on an underlying spring-embedding of agglomerated networks of all 12 cases. In this visual rendering of the lexicons, the size of a node (term) is proportional to its correlation centrality as defined in Equation (5). The individual lexicons, as shown in Figure 3, already show through visual inspection that they differ in size but show little systematicity, thereby defying simple classifications or categorizations. Visual representations, as shown in Figure 3, are suggestive, but quite unreliable for making inferences about differences. Therefore, we turn to more controllable ways to estimate the similarity of lexicons.
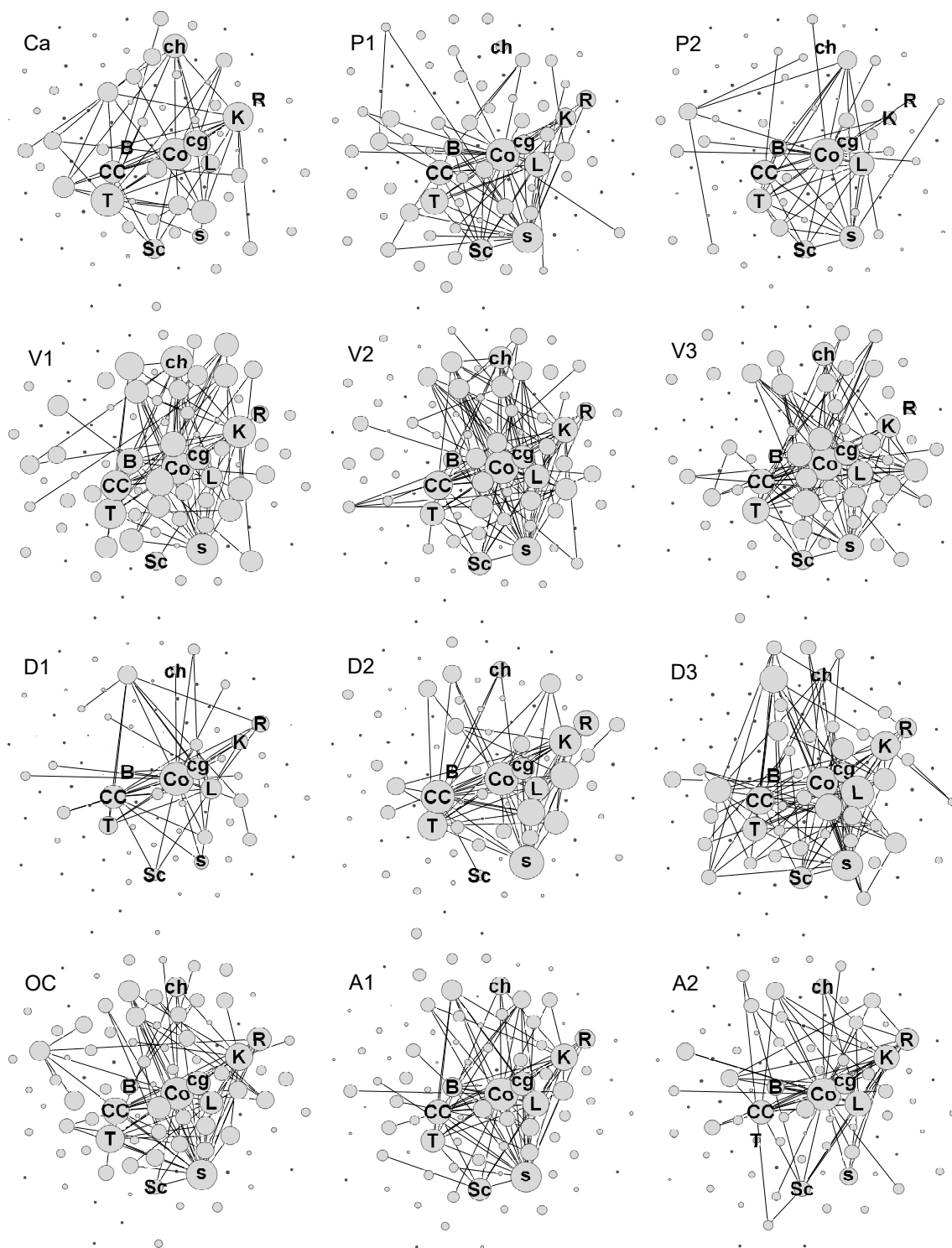
Figure 3. The lexicons of 12 research articles on conceptual change. The 12 lexicons visualized as spring-embedded networks. The acronyms of lexicons (upper left corner) identify the texts, as summarized in Table 1. Figure provides a schematic view, and therefore, only a few nodes are identified: Co = concept, CC = conceptual change, cg = change, ch = child, B = belief, K = knowledge, L = learn, R = research, s = student, and Sc = science. In each lexicon, a given node is always in the same position. The size of each node is proportional to its correlation centrality (as defined in Equation (5)).

3.3. Divergence and Similarity of Lexicons

Networks appear differently for different values of weight parameter β, and to decide when a stable region is reached and networks can be compared, we need to monitor the behavior of divergence when the value of β is increased. Examples of the behavior of Jensen–Shannon divergence J (based on concurrence Q) with increasing β are shown in Figure 4a–c (upper row) for lexicons V1, OC, and A1. In each case, the 11 different sets shown correspond to the results when a given lexicon is optimized for the best overlap with all the other 11 lexicons (denoted in the legends in upper row). In all cases, we can see that divergence grows rapidly in region $2 < \beta < 5$, reaches a maximum at $5 < \beta < 10$, and eventually reaches constant values for $\beta > 30$. This behavior results when increasing β, links with different weights contribute to the divergence. In regions with of low values of β, all links are weak, and since networks contain always same nodes, they are essentially disintegrated but similar. Gradually, with increasing values of β, many links begin to gain importance with increased weighting, and many of them begin to contribute to the similarity (or rather, dissimilarity) of networks; divergence reaches a maximum. With increasing weighting, however, only the most important links with the highest values of concurrence continue to contribute to the similarity and divergence decreases. Such a decrease in divergence indicates that lexicons possess, after all, quite a few term pairs connected by high concurrence links. Therefore, with the most important links dominating the behavior of the divergence at the highest values of β, divergences do not change anymore. In this region, one can reliably identify the key nodes (terms) and their rankings.
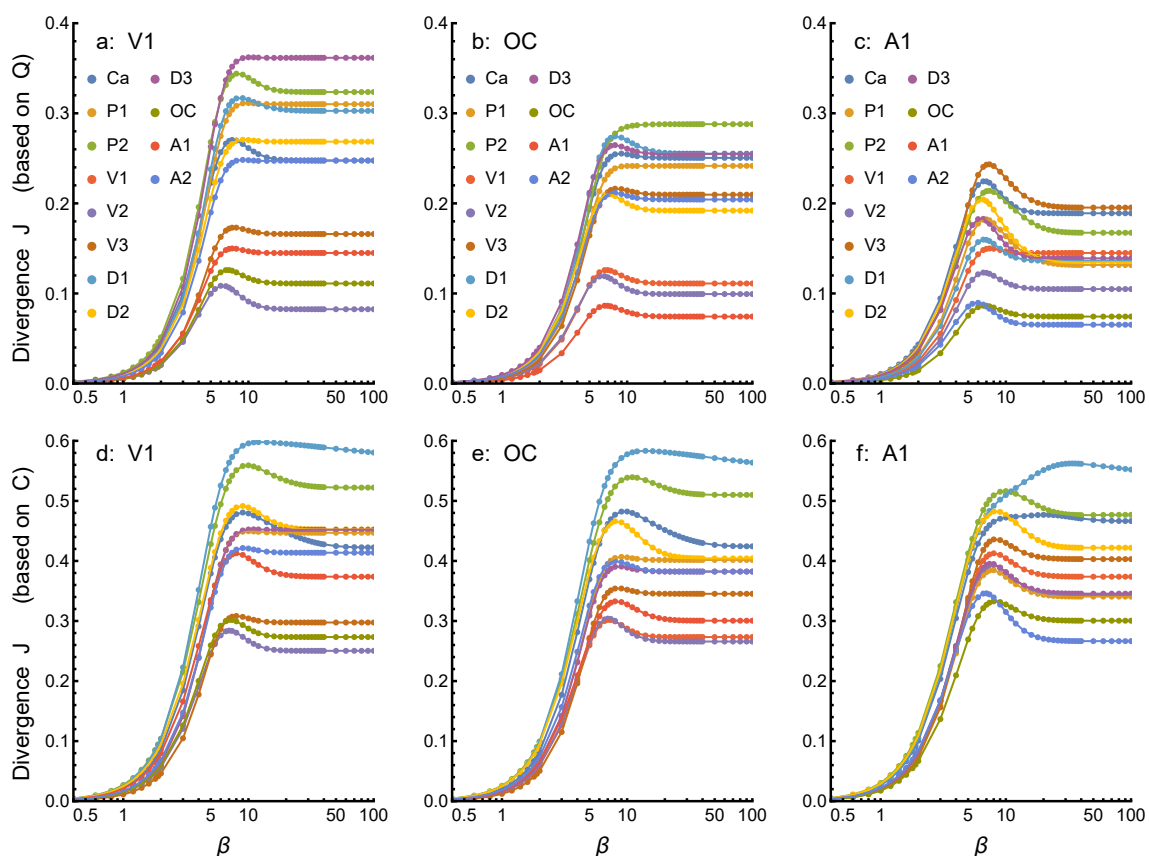

Figure 4. Jensen–Shannon divergences J for three lexicons, V1, OC, and A1. The divergence $J(A\|B)$ with increasing β is shown in panels (**a**–**c**) (upper row) for lexicons V1, OC, and A1, based on concurrence Q; and in panels (**d**–**f**) (lower row) based on contingency C. In each case, divergence J for a given lexicon A as compared to all other 11 lexicons B is as indicated by acronyms in legends.

The divergences J for optimized lexicons show large variation. Lexicons OC (Figure 4b) and A1 (Figure 4c) have much lower divergences in comparison to V1 (Figure 4a). Such

differences cannot be observed in the case of non-optimized lexicons corresponding to modified contingency C and shown in Figure 4d–f (lower row). The comparison of results in Figure 4a–c to those in Figure 4d–f (note the difference in scales) for Q- and C-based divergences demonstrates the important effect of optimization on matching the lexicons. However, the relative ordering of divergences is somewhat similar in both cases; optimization increases match, but does not dramatically change how relative differences behave.

Jensen–Shannon- divergence J (JSD) provides a basis to define the similarity S_{JSD} of lexicons, as introduced in Equation (7). The similarity S_{JSD} takes into account all links in the lexicons, and thus, is a global similarity measure. The cosine similarity S_{COS} in Equation (8), on the other hand, is based on the correlation centrality of nodes, i.e., off-diagonal sums of the correlation matrix. Figure 5 shows the similarities S_{JSD} and S_{COS} (upper and lower panels, respectively) for C- and Q-based lexicons (left and middle panels, respectively) as deviations from the median similarity (values shown above legends). The absolute change ΔS in similarity when C-based lexicons are optimized to Q-based lexicons is shown in Figure 5 in the right-hand panels.
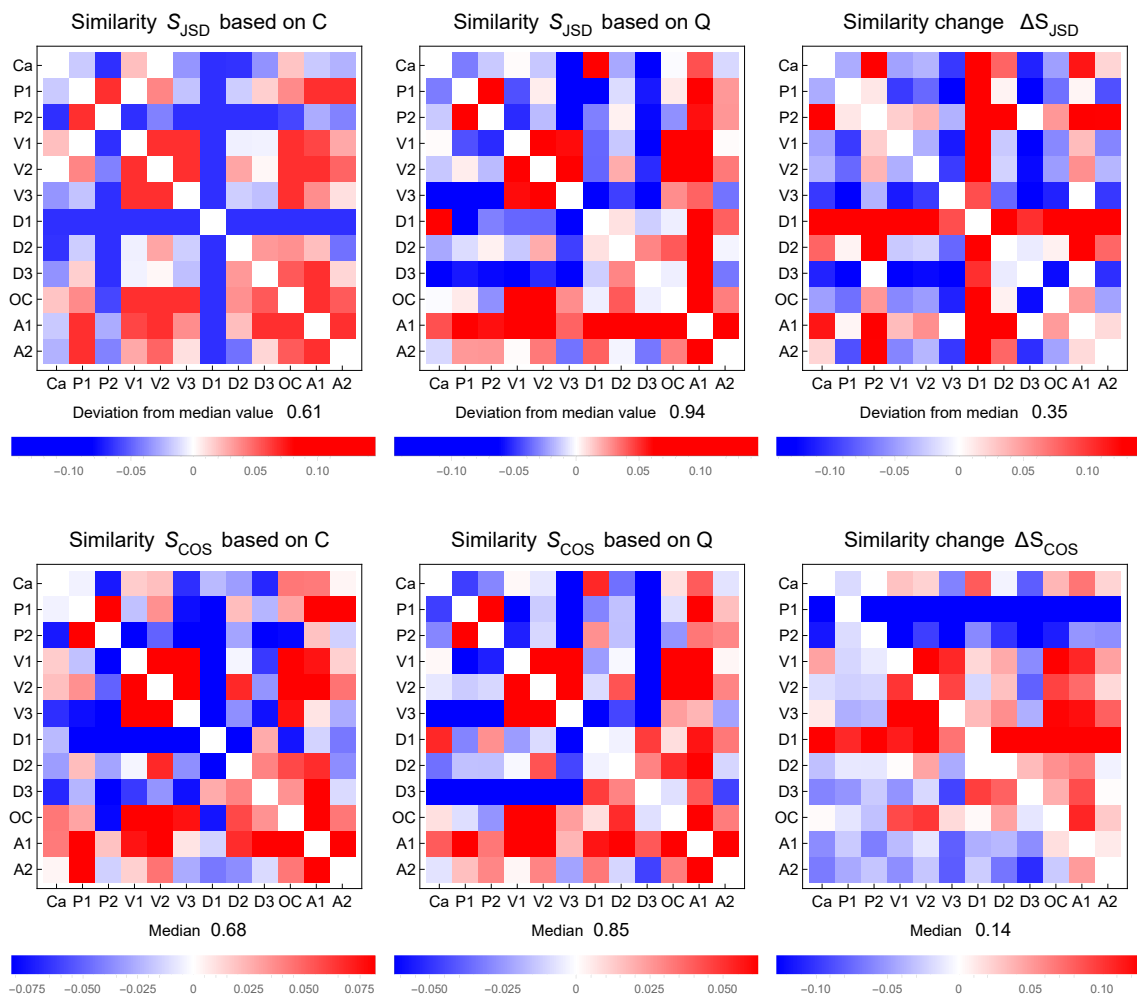

Figure 5. Similarity comparisons of lexicons. The similarity S_{JSD} based on Jensen–Shannon divergence is shown in the upper row for contingency C (at left) and concurrence Q (middle). The changes ΔS due to tuning are also shown (right). All values are shown as deviations from the median (value provided above bar-legend). The cosine similarity S_{COS} is shown in the lower row. In both cases, results are for a stabilized state with $\beta = 50$.

Results in Figure 5 for S_{JSD} show that in all cases the increase in similarity due to optimization is significant, although relative patterns as related to median values are somewhat similar, as already seen in three examples in Figure 4. However, some interesting

changes in relative patterns are notable. The clusters V1–V3; D1–D3; and OC, A1, and A3 are also present in contingency-based similarity patterns (upper row, left), but D1 is not strongly featured in the cluster D1–D3. It is also noteworthy that A1 in Q-based similarity is strongly similar to all other lexicons, but not to the same degree as in C-based similarity. However, in all cases all similarities increase significantly when lexicons are changed from C-based (median similarity 0.61) lexicons to Q-based ones (median similarity 0.94) for better matching.

It is interesting to note, as Figures 4 and 5 reveal, that texts V1, V2, and V3 are by the same lead authors, and thus supposedly belonging to the same thematic area and topics, so they are similar to each other. However, e.g., V1 is most similar to V2, yet it is more similar to OC and A1 than to V3. This indicates that V1 and V3 differ somewhat in their vocabularies, because apparently the authors wrote about their topic with differing emphasis, but nevertheless, review articles OC and A1 with the purpose of covering all aspects as thoroughly as possible, attained even larger similarities. In concordance with that notion, lexicon OC corresponding to a review article is most similar to V1 and V2, and next to them to A1, which is also the lexicon of a review article. Lexicon A1, on the other hand, is most similar to A2, OC, and V2, in that order. Contrary to these more or less expected findings, Figure 5 shows that sets D1, D2, and D3 of lexicons are less similar to each other than they are to lexicon A1; the authors of D1, D2, and D3 have to some degree different vocabularies in their texts, but the review article manages to cover all of them. In addition, V1, V2, and V3 all have quite low similarity to D1, D2, and D3, which was expected, because the main dividing line of thought is known to be between these disciplinary sub-groups. There are other more minor details of some interest, but for this study attempting to demonstrate the viability of the proposed method, it is not necessary to discuss those differences or their origins in more detail.

Scatter plots of similarities S_{JSD} and S_{COS} for all 12 lexicons are shown in Figure 6, based either on contingency C or concurrence Q (left and in the middle, respectively, in Figure 6). In scatter plots, Kendall's tau (non-parametric) correlation coefficients τ for the data points are also provided. The results in Figure 6a,b (upper row) are shown for $\beta = 0.3$ and for a stabilized state with $\beta = 50$ (Figure 6d,e in lower row). It is seen that at low values of β, the correlation between Q- and C-based cos-similarities S_{COS} is larger than for JSD similarities S_{JSD}. For comparison, straight lines indicate the reference lines of equal values, to show an increase in similarity owing to the optimization (i.e., all data points are well above the reference lines). As cos-similarity puts weight on key terms, the behavior shown in Figure 6 indicates that key terms contribute significantly to similarity. Lower values of correlation for JSD-based similarities in comparison to cos-similarity indicate larger variety in weak links corresponding to auxiliary terms. This is a conclusion that can also be inferred from the examples shown in Figure 2. The results in Figure 6c,f on the left display how cos- and JSD-similarities are correlated for C- and Q-based lexicons at low and high values of β. We can see that in the stabilized region, Q- and C-based similarities align, while in initial, low values of β, the scatter of values is much larger. The main message of this figure is that the best correlation of similarities is achieved between C- and Q-lexicons in stabilized region $\beta \gg 1$, meaning that in this stage, the key terms of optimized lexicons dominate the similarity. In this region, cos-similarity and JSD-similarity provide essentially the same results.

From the results in Figures 5 and 6, we can conclude that some lexicons of some texts are highly overlapping. For example, lexicons V1 and V2 are quite similar to each other, as are also lexicons P1 and P2 and A1 and A2. The texts that correspond to these lexicons had the same authors. In addition, lexicon A1 is highly similar to nearly all other lexicons, and OC also overlaps with many of the lexicons. These lexicons are from review-like texts, and A1 in particular attempts to cover many different views on conceptual change. In optimization, the similarity gain is also largest for the lexicons A1 and OC.
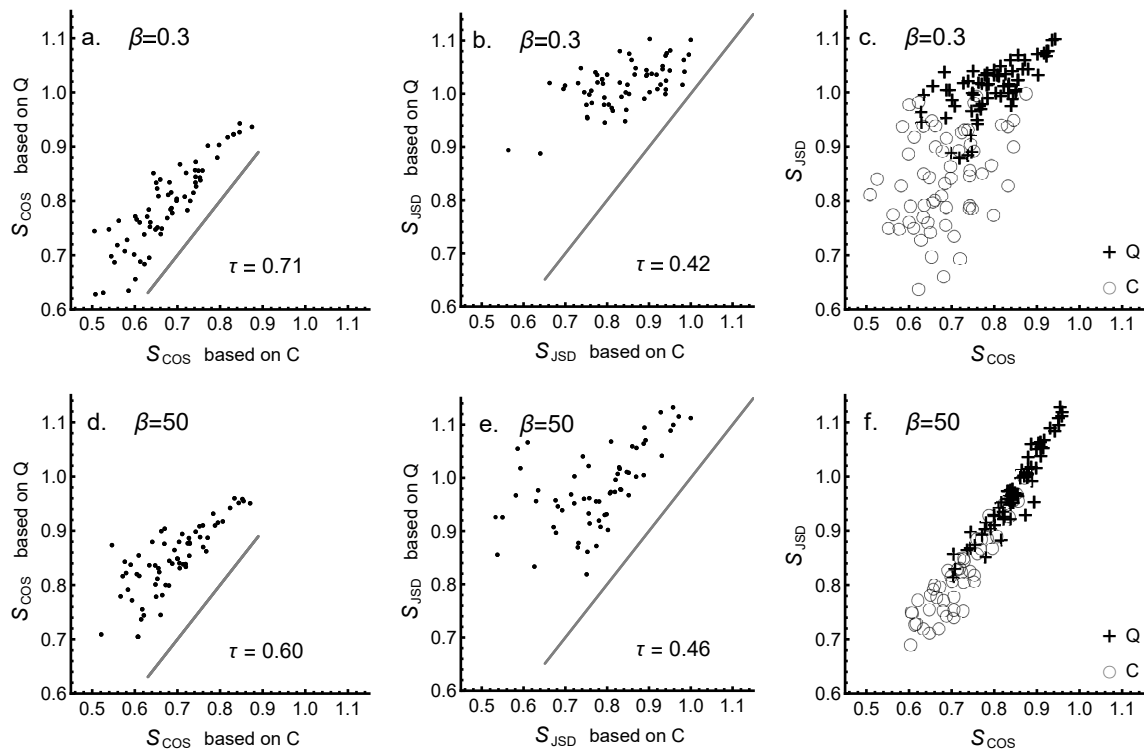
Figure 6. Scatter plots of similarities S_{JSD} and S_{COS} for all 12 lexicons. Similarities based on concurrence Q and contingency C (S_{COS} left) and S_{JSD} (in the middle). Straight lines indicate the references of equal value. Values of (non-parametric) Kendall's tau correlation coefficient τ are provided. The values corresponding to the initial state $\beta = 0.3$ are shown in the upper row, and those corresponding to the final stabilized state $\beta = 50$ in the lower row. The right panel shows scatter plots of S_{JSD} against S_{COS}.

Finally, it is of interest to compare the similarity of lexicons based on co-occurrence counts to the similarity obtained on the basis of frequency distributions shown in Figure 1. Such a comparison can be carried out using JSD-similarity, which can be calculated from Equations (6) and (7) for ordinary probability distributions by replacing density matrices with probability density distributions and traces with sums. The resulting pairwise similarities of lexicons are similar to results shown in Figures 5 and 6, and again the same high similarity clusters, as discussed above, can be discerned. For comparison, the Kendall–tau correlations of frequency-based lexicon JSD-similarities to concurrence and contingency-based similarities are now 0.37 and 0.35, respectively. The differences are not large, indicating that few frequently appearing terms are also the terms that connect to other frequently occurring terms. Next, we turn to identifying the key terms responsible for the similarity and how they are shared in different lexicons.

3.4. Key Terms and Shared Key Terms

A rough idea of the rankings of key terms is provided in Table 4, where the top-most key terms are listed in cumulative order, for C- and Q-based lexicons and for low and high values of β. The key terms are listed in four cohorts. Cohort I contains all terms that are found among the top 10 terms in at least eight lexicons, cohort II contains terms found among the top 20 but not among cohort I, and similarly cohorts III and IV contain terms that appear in the top 40 and 60 but not in any of the lower cohorts. From the results in Table 4, we can see that in C-based lexicons, cohort I contains only a single shared term, in contrast to Q-based lexicons, where several terms are shared.

Table 4. Cumulative appearance of terms in cohorts of top 10 (I), 20 (II), 40 (III), and 60 (IV) terms, for concurrence (Q) and contingency (C)-based ranking. Terms that appear in at least 8 lexicons are listed. In the column for $\beta = 0.3$, terms that also appear for $\beta = 50.0$ are in boldface, and in the corresponding column, only additional terms are listed.

Cohort	$\beta = 0.3$	$\beta = 50$
Q-I	**concept, conceptual change, knowledge, student**	theory/ies
Q-II	**change, learn**	process
Q-III	**belief, children, explanation, model,**	
	research, science, term, process	nature, idea
Q-IV	**construct, epistemol., experience**	naive
	information, misconception, p-prim, world	
	nature, idea, analog, coherence	
C-I	**concept,** student	
C-II	**change, learn,conceptual change, knowledge, thery/(ies)**	student
C-III	**belief, children,research, science, term,**	
	process,nature, explanation, idea	
C-IV	**construct, epistemol., experience**	naive,idea
	information, misconception, p-prim, world	explanation
	analog, coherence, evidence, model	

This demonstrates the significance of optimization. In cohorts II–IV we can see many differences when *C*- and *Q*-cohorts are compared. Although the collections of terms are nearly the same, their positions within cohorts changed. Terms that appear in higher Q-cohorts in comparison to their positions in C-cohorts are the terms that need to increase their value of concurrence with other terms due to optimization (e.g., terms "conceptual", "change", "knowledge", and "theory") to become better shared, and very few terms (e.g., "nature") were moved lower cohorts due to optimization. The rankings of top terms in *C*- and *Q*-based lexicons are thus different, although the differences are not dramatic. It is noteworthy that in both *C*- and *Q*-lexicons, the effect of β on the rankings of the upper-most terms was not very significant. This is in agreement with the result that cos-similarity (which weights key terms) did not change dramatically with increasing value of β (see Figures 5 and 6).

Taking the *Q*-based correlation centrality as the basis of the rankings, the occurrences of the top-ranking terms in different lexicons are shown in Figure 7. The first panel, Figure 7a, shows how frequently a given term (listed on the right) occurs among the top 10 terms. If, in a given lexicon, the term is found among the top 10 ranking terms in all 11 optimized lexicons, the symbol is dark purple; the fewer the occurrences, the lighter the blue symbol. The occurrences among the 20 and 40 top ranking terms are shown in panels Figure 7b,c, respectively; and finally, in panel Figure 7d, for the 60 top ranking terms. Only cases where a given term is found in eight lexicons are shown.

In all texts, "concept", "conceptual change", "change(e)", "student", and "learn(ing)" are among the top 10 terms; and "knowledge" and "theori(es)" are among the top 10 in most lexicons. Other terms occur more sporadically, and one has large variability between lexicons in the distribution of top terms. When attention is based on occurrence among the 20 top terms, new frequently-occurring terms emerge. Among them are "science" and "research(ing)". Furthermore, "model" can be found among the top 20 terms in many lexicons. In the set of 40 top ranking terms, new frequently-occurring terms are "structure", "coherence", "belief", and "misconception". This shows how different common themes emerge gradually as expanding shared sets of terms; the lexicons are overlapping. Finally, in the set of the 60 highest ranking terms, two new blocks of interest emerge. One block corresponds to framework theories with terms "frameworktheori(es)" and "coher(ence)"; the other block consists of knowledge-as-elements view terms "knowledge(in)piec(es)" and

"p-prim". In addition, the terms "process", "exper(ience)", and "explanation" are featured among the top 60 terms.
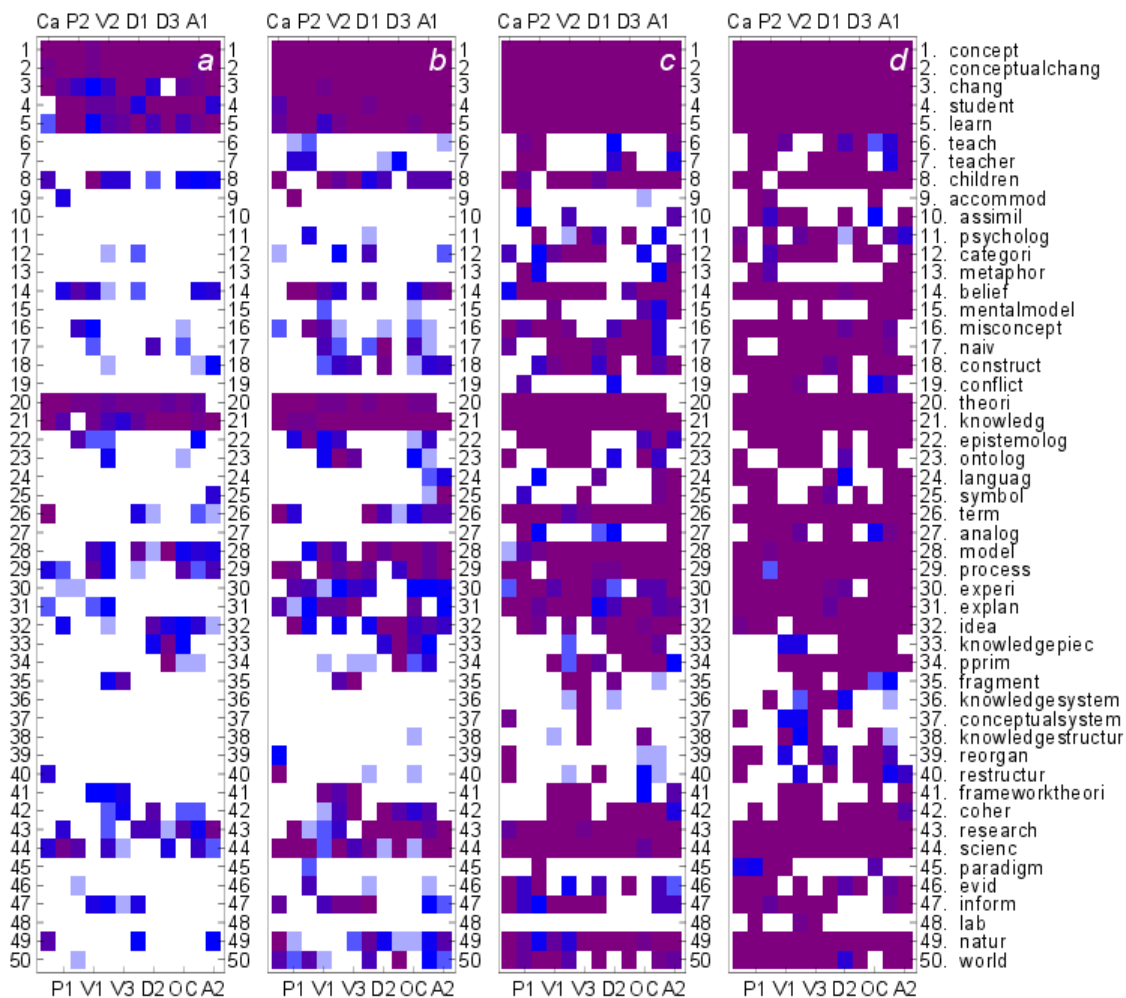

Figure 7. Patterns of shared terms in optimized lexicons based on concurrence Q. Occurrences of terms in top 10 (**a**), 20 (**b**) 40 (**c**), and 60 (**d**) cohorts in all 12 cases in the stabilized case with $\beta = 50$. When a term in the lexicon occurs in all 11 tuned cases, the symbol is purple; otherwise blue. The lighter the symbol, indicating occurrence, the fewer the tuned lexicons where the term occurs. Note that only the terms occurring at least eight times are shown.

The breakdown of terms in cohorts of 10, 20, 40, and 60 top-ranking cohorts already gives us an idea how terms and their semantic fields are shared by different optimized lexicons, leading to the overlap and similarity of the lexicons. The top-ranking terms listed in Figure 7 are also the terms denoted by large nodes in Figure 3. Together, the information provided by these representations establishes a qualitative idea of which lexicons are similar and on what basis of shared terms.

The breakdown of key terms in top-cohorts as based on concurrence Q can be compared with similar breakdown when C-based correlation centrality is used as the basis of ranking. The breakdown in top cohorts in that case is shown in Figure 8. Comparing the results in Figure 7 to results in Figure 8, we see that with contingency C, key terms are not shared as frequently as for concurrence Q; many lexicons that overlap by sharing key terms with Q drop out for C. Comparison in Figures 7 and 8 finally answers the question of which terms are the significant, tunable terms that are responsible for the increased similarity of lexicons; they are the terms appearing in Figure 7 but not in Figure 8. In cases of cohorts of top 10 and 20, we can locate several such important, tunable terms. On the other hand, the

terms that appear in Figures 7 and 8 both, i.e., in practice nearly all terms in Figure 8, are not tunable, or their roles are not changed by tuning. Therefore, in seeking better overlap of lexicons, terms appearing in cohorts of top 10 and 20 terms in Figure 7 but not in Figure 8 are the most important ones.
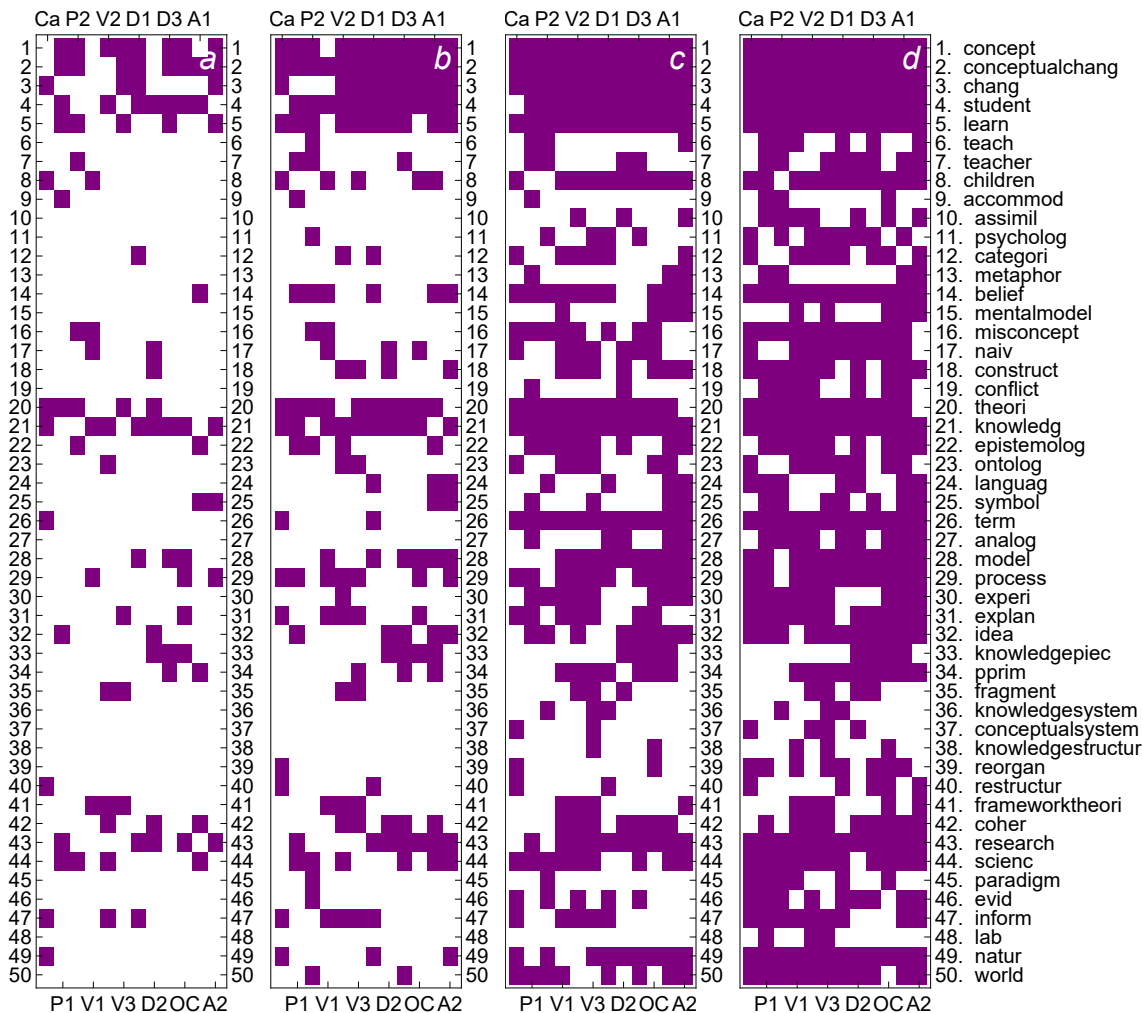

Figure 8. Patterns of shared terms in lexicons based on concurrence *C*. Occurrences of terms in top 10 (**a**), 20 (**b**) 40 (**c**), and 60 (**d**) cohorts in all 12 cases in the stabilized case with $\beta = 50$. Terms occurring at least eight times are shown.

4. Conclusions

In this study, we have introduced an approach to constructing and comparing lexicons of terms in scholarly texts. The purpose of the method is to provide a relatively simple means to compare disciplinary differences as they appear in texts, and moreover, to study the extent to which lexicons can be optimized for the best possible overlap. The approach suggested here consists of three steps, each of which warrants some brief discussion.

First, we have adopted and utilized the recently suggested quantum semantics [38–40] to construct and optimize lexicons. Quantum semantics allows us to describe how subjective biases affect connections between words, and thus, give the possibility of different readings of the meanings of words. According to quantum semantics, the meanings of words and terms are entangled so that different meanings can not be entirely separated [38–40]. The "quantumness" in this context does not refer to any physical-like quantization, but rather, to a specific kind of a logic behind quantum physics, which allows the superposition and entanglement of states (here, co-occurrence of words). Such a logical structure is provided by quantum logic, and therefore, the approach borrowing it is referred to as quantum

semantics. A major shortcoming of quantum semantics as applied to a written text is the ambiguity that remains in inferring the phase factors (see Appendix A) that determine the degree of entanglement [38–40]. Due to that limitation, compound phase Θ remains a fitting parameter. Although in principle the individual phase factors can be obtained (see reference [39] for a discussion), it is impossible with the sample and data in focus here. Therefore, we have only discussed results corresponding to extreme cases with classical contingency C (no optimization) and complete optimization of concurrence Q. This shows the possible bounds of variation of achieving overlap in lexicons, and thus, possibilities for improved agreement of meaning as far as it is reflected in similar co-occurrence of words in expressions. Despite many limitations, quantum semantics thus provides an alternative to approaches in which semantic fields are constrained to have fixed meanings, without the possibility of subjective bias. Interestingly, recent advances in quantum semantics [81] may open a way to discuss the role of subjective experiences (e.g., including affective factors) in determining the phase factors, and thus, tuning of lexicons. This possibility aligns well with approaches where sentiment analysis is included as part of analysis of cognitive networks (see, e.g., reference [31]. However, further discussions of such extensions is beyond the scope of present study.

Second, on the basis of quantum semantics, lexicons of terms were constructed in the form of networks and analyzed by utilizing path (walk) counting in terms of exponential matrix transformation, which is widely used as the basis of network analysis [42–45]. Based on exponential matrix transformation, a density matrix and correlation matrix [47] were introduced to characterize the lexicons. The correlation matrix allowed us to define the correlation centrality of nodes, used here to identify the important nodes in the lexicon.

Third, and finally, by using density matrices, comparisons of lexicons could be based on information theoretic relative entropy (Jensen–Shannon divergence, JSD) [48,50–53]. Apart from the use of a correlation matrix, as suggested in reference [47], the other steps were standard ones from other similarly focused studies, thereby requiring no further discussion.

The method suggested here was tested by applying it in the analysis of 12 scholarly texts about research in conceptual change. The results show that the method is able to bring out the disciplinary differences, in agreement with the expectations and known differences in the positions taken by different authors. That finding in itself was of course not unexpected, and many features revealed by the semantic analysis were probably easy to anticipate for those familiar with the content of the papers (see, e.g., reference [54] and references therein). It was also found that although concurrence, contingency, and frequency-based similarities produce different results, the differences are not dramatic; in all cases nearly the same clusters of texts turned out to be the most similar to each other. This indicates that frequently occurring terms are also the terms responsible for the strongest links between term pairs. Such a dependence is expected on the basis of tight focus of texts, and furthermore, a certain idiosyncratic way of authors discussing the topics.

Nevertheless, the results suggest that overlap of lexicons can be improved by optimization (i.e., tuning for best agreement). If one takes the lexicons central for communicating ideas across different disciplinary groups, as in Kuhn's view [16–18], the ability to improve overlap can be taken as a proxy to enhance and improve communication of ideas and find a consensus of views.

In summary, we have demonstrated that a relatively simple (but many stepped) method of analysis which is a generalization of word co-occurrence counting is a viable way to construct and compare lexicons. It brings out the disciplinary differences in scholarly texts, and thus can be used to study differences in how various disciplinary schools frame their key concepts and use them in disseminating their ideas. We believe that having a sufficiently simple method for such analyses will help to lower the threshold to adopt network-based methods in fields of research (e.g., educational and behavioral science) where such methods are not yet well known, but where they might provide significant support for more traditional interpretative research approaches.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable

Data Availability Statement: Not applicable.

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

Appendix A. Concurrence

In quantum semantics [38–40], word co-occurrence is described as a two-state event: words A and B may both occur at least once in a block of text, only one of them might occur (at least once), or neither A nor B occurs. Occurrence is described by tag 1 and non-occurrence by 0. By adopting this simple prescription for co-occurrence counting, the outcomes can be coded as pairs $|00>$ (neither A nor B occur), $|01>$ (A does not occur but B occurs), $|10>$ (A occurs but not B), and $|11>$ (A and B both occur). These states $|00\rangle$, $|11\rangle$, $|01\rangle$, and $|10\rangle$ are the basic building blocks of the well-known qubit states of two two-state systems [38–40]. Note that we use here the so called bra-ket notation for states, but this is not essential here, and notation can be taken as a shorthand for indexing states of co-occurrence. The frequencies to find each of them in counting the co-occurrence of words in sentences of a given text are denoted by n_{00}, n_{01}, n_{10}, and n_{11}, respectively.

Having defined the elementary states $|00\rangle$, $|01\rangle$, $|10\rangle$, and $|11\rangle$, we can construct a superposition of these states, corresponding to the most general possibility of all potential ways to build up the word meaning as it is realized in different sentences. Such a state can be constructed from a two qubit pure states in the form [38,40]

$$|\Psi\rangle = c_{00}|00\rangle + c_{01}|01\rangle + c_{10}|10\rangle + c_{11}|11\rangle \tag{A1}$$

where amplitudes c_{ij} are complex numbers related to real valued frequency n_{ij} with normalization $\Sigma n_{ij} = \Sigma |c_{ij}|^2 = 1$ and also including phase factors ϕ_{ij} in the form

$$c_{ij} = \sqrt{n_{ij}} \, e^{i\phi_{ij}} \tag{A2}$$

As discussed in detail in reference [38], phase factors ϕ_{ij} can be taken as parameters that describe various possible subjective ways to read meanings into the co-occurrence of terms A and B, i.e., to emphasize the meaning of a connection differently. Such biased and subjective ways to emphasize connections are supposedly at least one factor behind the differences in understanding meanings of words [38–40], and with similar arguments, can be also extended to cases where scientific terms have different, disciplinary-dependent meanings in different scholarly texts. The question of how phase factors should be adjusted, and how they are related in detail in subjective factors, is an unsolved problem, and therefore, they are usually treated as post factum fitting parameters (see reference [39] for a detailed discussion). Nevertheless, the expressions in Equations (A1) and (A2) now allow us to take into account a kind of entanglement effect due to biased emphasis on co-occurrence.

Factorization of state $|\Psi\rangle$ in Equation (A1) into independent marginal states of single qubits is not possible in general; the state is said to be entangled. Different entanglements then correspond to different subjective, biased readings of the meaning of co-occurrence. Recently, it has been suggested that the degree of entanglement can be quantified as a difference between the factorized state and entangled state, called a concurrence Q [38,40], following the way the entangled states are characterized in quantum information theory [68]. The basic idea behind such an approach is a notion that entanglement becomes measurable by switching the order of appearance of the co-occurrence, thereby changing the state as opposite to the original [40]. Technically, as pointed out by Galofaro et al., this

is an "abstract machine" corresponding a quantum logic gate as applied to word meaning [40]. Such an "abstract machine" of transformation of meaning can be now constructed by utilizing Pauli spin-operator σ_y, with the result [38,68]

$$Q = \langle \Psi | \, \sigma_y \otimes \sigma_y \, | \Psi^* \rangle \tag{A3}$$

where $*$ denotes a complex conjugate. Evaluating the expression in Equation (A3) for state in Equation (A1) is straightforward, leading to the result [38]

$$Q = 2 \, | \, c_{01} c_{10} - c_{00} c_{11} \, | , \; \; 0 \leq Q \leq 1 \tag{A4}$$

Next, by substituting in Equation (A3) the amplitudes in Equation (A2), we finally get the concurrence in terms of co-occurrence frequencies and phase factors, in the form [38]

$$Q = Q_0 \sqrt{ 1 - \Theta \frac{ \sqrt{ n_{01} n_{10} n_{00} n_{11} } }{ (n_{01} n_{10} + n_{00} n_{11}) / 2 } } \, , \tag{A5}$$

where $Q_0 = 2\sqrt{ n_{01} n_{10} + n_{00} n_{11} }$ and $\Theta = \cos(\phi_{01} + \phi_{10} - \phi_{00} - \phi_{11})$ is the compound phase factor, taken here as the free parameter to account for (unknown) phase factors. It is interesting to note that the last term in the square root is the ratio R of the geometric mean of $n_{10} n_{01}$ and $n_{00} n_{11}$ to their arithmetic mean. With these abbreviations concurrence, one obtains a simple form $Q = Q_0 \sqrt{ (1 - \Theta R) }$ of Equation (1), where entanglement is now possible only in cases $R \neq 0$.

Appendix B. Density Matrix and Correlation Matrix

In a very basic level, a network is characterized by its nodes (vertices), links between nodes (edges) and weights of the links, and the corresponding relevant magnitude or value characterizing the connection. This information is contained in a condensed form in weighted adjacency matrix $\mathbf{W}$, where entries $[\mathbf{W}]_{ij} = W_{ij}$ provide the weight of a link between nodes i and j if there is a link connecting the nodes; otherwise, $W_{ij} = 0$. Here, we assume that matrices are symmetric, i.e., $W_{ij} = W_{ji}$.

A probability density-like quantity called density matrix $\boldsymbol{\rho}$ characterizing the network can be introduced through matrix exponential transformation $\exp[\beta \mathbf{W}]$ [42–46]. This transformation is related to weighted path (or walk) counting between nodes, as can be seen through power-series expansion

$$\exp[\beta \, \mathbf{W}] = \mathbf{I} + \frac{\beta^1 \mathbf{W}^1}{1!} + \frac{\beta^2 \mathbf{W}^2}{2!} + \frac{\beta^3 \mathbf{W}^3}{3!} \dots , \tag{A6}$$

where matrix power $\mathbf{W}^k$ counts weight of paths (walks) of length k, and each element $\exp[\beta \mathbf{W}]_{ij}$ provides the weighted count connecting nodes i and j. Note that the exponential transformation weights the paths according to their length k by a weight factor β^k and the inverse of the factorial $k!$. Therefore, connections with short paths gain more importance than those with long paths.

When we normalize the exponential transformation in Equation (A6), we obtain the density matrix

$$\boldsymbol{\rho} = Z^{-1} \exp[\beta \mathbf{W}] , \tag{A7}$$

where $Z = \mathrm{Tr} \exp[\beta \mathbf{W}]$ is a normalization factor. Due to normalization, $\mathrm{Tr} \boldsymbol{\rho} = 1$ the matrix has an analogous role to probability density; due to its constructions based on a real and symmetric matrix $\mathbf{W}$, it is Hermitian; and finally, due to exponential matrix transformation, it is a positive semidefinite matrix; therefore, it can be taken as a density matrix [42–44,47].

The density matrix $\boldsymbol{\rho}$ provides a way to formulate a correlation matrix describing how nodes' positions in the network as they relate to their adjacency (i.e., steps separating the nodes) are correlated [47]. Following Estrada's derivation (see reference [47]), we introduce a matrix $\boldsymbol{\Gamma}$ in the form

$$\mathbf{\Gamma} = \gamma^{-\frac{1}{2}} \boldsymbol{\rho} \, \gamma^{-\frac{1}{2}} \,, \tag{A8}$$

where $\gamma = \mathrm{Diag}[\mathbf{\Gamma}]$ is a diagonal matrix. Note that quantity $\mathbf{\Gamma}$ in Equation (A8) differs from definition of similar quantity *C* in reference [47] (Equation 3.3) only by normalization factor *Z* we use to define density matrix in Equation B.2, but it cancels out, making Equation (A8) equivalent to Equation (3.3) in reference [47]. To avoid confusion, we remind the reader that for obvious reasons of notation used in the present study, we have used different symbols in comparison to reference [47].

By tracing the steps of argument in reference [47] (from Equation 3.3. to 3.6), the elements of $[\mathbf{\Gamma}]_{ij}$ of the correlation matrix can be shown to be directly related to the covariance of values of nodes *i* and *j* in the network [47]. Here, these intermediate steps or arguments are not repeated, because they are exactly the same as in reference [47]. The correlation matrix Γ provides now a way to find the key nodes on the basis of correlations; nodes correlated strongly with other nodes are important, and the weaker correlation, the less the significance of the node.

References

1. Fanelli, D.; Glänzel, W. Bibliometric evidence for a hierarchy of the sciences. *PLoS ONE* **2013**, *8*, e66938. [CrossRef] [PubMed]
2. Fanelli, D. "Positive" results increase down the hierarchy of the sciences. *PLoS ONE* **2010**, *5*, e10068. [CrossRef] [PubMed]
3. Kuhn, T.S. *The Structure of Scientific Revolutions*, 1st ed.; University of Chicago Press: Chicago, IL, USA, 1962.
4. De Solla Price, D.J. The Science of Science. *Bull. At. Sci.* **1965**, *21*, 2–8. [CrossRef]
5. Garfield, E. Citation indexes for science: A new dimension in documentation through association of ideas. *Science* **1955**, *122*, 108–111. [CrossRef]
6. Klavans, R.; Boyack, K.W. Toward a consensus map of science. *J. Am. Soc. Inf. Sci. Tech.* **2009**, *60*, 455–476. [CrossRef]
7. Leydesdorff, L.; Rafols, I. Interactive overlays: A new method for generating global journal maps from Web-of-Science data. *J. Informetr.* **2012**, *6*, 318–332. [CrossRef]
8. Leydesdorff, L.; Carley, S.; Rafols, I. Global maps of science based on the new Web-of-Science categories. *Scientometrics* **2013**, *94*, 589–593. [CrossRef]
9. Zeng, A.; Shen, Z.; Zhou, J.; Wu, J.; Fan, Y.; Wang, Y.; Stanley, H.E. The science of science: From the perspective of complex systems. *Phys. Rep.* **2017**, *714–715*, 1–73.
10. Fortunato, S.; Bergstrom, C.T.; Börner, K.; Evans, J.A.; Helbing, D.; Milojević, S.; Petersen, A.M.; Radicchi, F.; Sinatra, R.; Uzzi, B.; et al. Science of science. *Science* **2018**, *359*, e0185. [CrossRef]
11. De Domenico, M.; Omodei, E.; Arenas, A. Quantifying the diaspora of knowledge in the last century. *Appl. Netw. Sci.* **2016**, *1*, 15. [CrossRef]
12. Ding, J.; Ahlgren, P.; Yang, L.; Yue, T. Disciplinary structures in Nature, Science and PNAS: Journal and country levels. *Scientometrics* **2018**, *116*, 1817–1852. [CrossRef]
13. Herrera, M.; Roberts, D.C.; Gulbahce, N. Mapping the Evolution of Scientific Fields. *PLoS ONE* **2010**, *5*, e10355. [CrossRef]
14. Milojevic, S. Quantifying the cognitive extent of science. *J. Informetr.* **2015**, *9*, 962–973. [CrossRef]
15. Leydesdorff, L.; Adina Nerghes, A. Co-word Maps and Topic Modeling: A Comparison Using Small and Medium-Sized Corpora. *J. Assoc. Inform. Sci. Tech.* **2017**, *68*, 1024–1035. [CrossRef]
16. Kuhn, T.S. *The Road since Structure*; University of Chicago Press: Chicago, IL, USA, 2000.
17. Hoyningen-Huene, P. *Reconstructing Scientific Revolutions*; University of Chicago Press: Chicago, IL, USA, 1993.
18. Gattei, S. *Thomas Kuhn's Linguistic Turn and the Legacy of Logical Empiricism*; Routledge: Oxon, UK, 2016.
19. Callon, M.; Courtial, J.-P.; Turner, W.A.; Bauin, S. From translations to problematic networks: An introduction to co-word analysis. *Soc. Sci. Inform.* **1983**, *22*, 191–235. [CrossRef]
20. Leydesdorff, L. Words and co-words as indicators of intellectual organization. *Res. Policy* **1989**, *18*, 209–223. [CrossRef]
21. Tijssen, R.; Van Raan, A. Mapping co-word structures: A comparison of multidimensional scaling and LEXIMAPPE. *Scientometrics* **1989**, *15*, 283–295. [CrossRef]
22. Lehmann, F. Semantic networks. *Comp. Math. Apll.* **1992**, *23*, 1–50. [CrossRef]
23. Carley, K.M. Network text analysis: The network position of concepts. In *Text Analysis for the Social Sciences: Methods for Drawing Statistical Inferences from Texts and Transcripts*; Roberts, C.W., Ed.; Lawrence Erlbaum: Mahwah, NJ, USA, 1997; pp. 79–102.
24. Leydesdorff, L.; Welbers, K. The semantic mapping of words and co-words in context. *J. Informetr.* **2011**, *5*, 469–475. [CrossRef]
25. Blei, D.M.; Lafferty, J.D. Topic models. In *Text Mining: Classification, Clustering, and Applications*; Srivastava, A.N., Sahami, A., Eds.; Taylor and Francis: London, UK, 2009; pp. 71–94.
26. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
27. Amancio, D.R.; Oliveira, O.N., Jr.; da Fontoura Costa, L. Structure–semantics interplay in complex networks and its effects on the predictability of similarity in texts. *Physica A* **2012**, *391*, 4406–4419. [CrossRef]

28. Kulig, A.; Drozdz, S.; Kwapien, J.; Oswiecimka, P. Modeling the average shortest-path length in growth of word-adjacency networks. *Phys. Rev. E* **2015**, *91*, 032810. [CrossRef] [PubMed]

29. Yin, X.; Wang, H.; Yin, P.; Zhu, H.; Zhang, Z. A co-occurrence based approach of automatic keyword expansion using mass diffusion. *Scientometrics* **2020**, *124*, 1885–1905. [CrossRef]

30. Zhu, X.; Zhang, Y. Co-word analysis method based on metapath of subject knowledge network. *Scientometrics* **2020**, *123*, 753–766. [CrossRef]

31. Teixeira, A.S.; Talaga, S.; Swanson, T.J.; Stella, M. Revealing semantic and emotional structure of suicide notes with cognitive network science. *Sci. Rep.* **2021**, *11*, 19423. [CrossRef] [PubMed]

32. Christianson, N.H.; Sizemore, B.A.; Bassett, D.S. Architecture and evolution of semantic networks in mathematics texts. *Proc. R. Soc. A* **2020**, *476*, 20190741. [CrossRef]

33. Chai, L.R.; Zhou, D.; Bassett, D.S. Evolution of semantic networks in biomedical texts. *J. Complex Netw.* **2020**, *8*, cnz023. [CrossRef]

34. Ribeiro, E.; Teixeira, A.S.; Ribeiro, R.; de Matos, D.M. Semantic frame induction through the detection of communities of verbs and their arguments. *Appl. Netw. Sci.* **2020**, *5*, 69. [CrossRef]

35. Medeuov, D.; Roth, C.; Puzyreva, K.; Basov, N. Appraising discrepancies and similarities in semantic networks using concept-centered subnetworks. *Appl. Netw. Sci.* **2021**, *6*, 66. [CrossRef]

36. de Arruda, H.F.; Silva, F.N.; Marinho, V.Q.; Amancio, D.R.; da Fontoura Costa, L. Representation of texts as complex networks: A mesoscopic approach. *J. Complex Netw.* **2018**, *6*, 125–144. [CrossRef]

37. Amancio, D.R. Probing the Topological Properties of Complex Networks Modeling Short Written Texts. *PLoS ONE* **2015**, *10*, e0118394. [CrossRef] [PubMed]

38. Surov, I.A.; Semenenko, E.; Platonov, A.V.; Bessmertny, I.A.; Galofaro, F.; Toffano, Z.; Khrennikov, A.Y.; Alodjants, A.P. Quantum semantics of text perception. *Sci. Rep.* **2021**, *11*, 4193. [CrossRef]

39. Surov, I.A. Quantum Cognitive Triad: Semantic Geometry of Context Representation. *Found. Sci.* **2021** *26*, 947–975. [CrossRef]

40. Galofaro, F.; Toffano, Z.; Doan, B.-L. A quantum-based semiotic model for textual semantics. *Kybernetes* **2018**, *47*, 307–320. [CrossRef]

41. Busemeyer, J.R.; Bruza, P.D. *Quantum Models of Cognition and Decision*; Cambridge University Press: Cambridge, UK, 2012.

42. Estrada, E. *The Structure of Complex Networks: Theory and Applications*; Oxford University Press: Oxford, UK, 2012.

43. Estrada, E.; Hatano, N.; Benzi, M. The physics of communicability in complex networks. *Phys. Rep.* **2012**, *514*, 89–119. [CrossRef]

44. Benzi, M.; Estrada, E.; Klymko, C. Ranking hubs and authorities using matrix functions. *Linear Algebra Appl.* **2013**, *438*, 2447–2474. [CrossRef]

45. Kunegis, J.; Fay, D.; Bauckhage, C. Spectral evolution in dynamic networks. *Knowl. Inf. Syst.* **2013**, *37*, 1–36. [CrossRef]

46. Dehmer, M. Information processing in complex networks: Graph entropy and information functionals. *Appl. Math. Comput.* **2008**, *201*, 82–94. [CrossRef]

47. Estrada, E. Informational cost and networks navigability. *Appl. Math. Comp.* **2021**, *397*, 125914. [CrossRef]

48. Biamonte, J.; Faccin, M.; De Domenico, M. Complex networks from classic to quantum. *Comm. Phys.* **2019**, *2*, 53. [CrossRef]

49. Faccin, M.; Johnson, T.; Biamonte, J.; Kais, S.; Migdał, P. Degree Distribution in Quantum Walks on Complex Networks. *Phys. Rev. X* **2013**, *3*, 041007. [CrossRef]

50. De Domenico, M.; Biamonte, J. Spectral Entropies as Information-Theoretic Tools for Complex Network Comparison. *Phys. Rev. X* **2016**, *6*, 041062. [CrossRef]

51. Gerlach, M.; Font-Clos, F.; Altmann, E.G. Similarity of Symbol Frequency Distributions with Heavy Tails. *Phys. Rev. X* **2016**, *6*, 021009. [CrossRef]

52. Ré, M.A.; Azad, R.K. Generalization of Entropy Based Divergence Measures for Symbolic Sequence Analysis. *PLoS ONE* **2014**, *9*, e93532. [CrossRef] [PubMed]

53. Dias, L.; Gerlach, M.; Scharloth, J.; Altmann, E.G. Using text analysis to quantify the similarity and evolution of scientific disciplines. *R. Soc. Open Sci.* **2018**, *5*, 171545. [CrossRef]

54. Potvin, P.; Nenciovici, L.; Malenfant-Robichaud, G.; Thibault, F.; Sy, O.; Mahhou, M.A.; Bernard, A.; Allaire-Duquette, G.; Sarrasin, J.B.; Brault Foisy, L.-M.; et al. Models of conceptual change in science learning: Establishing an exhaustive inventory based on support given by articles published in major journals. *Stud. Sci. Educ.* **2020**, *56*, 157–211. [CrossRef]

55. Carey, S. Reorganization of knowledge in the course of acquisition. In *Ontogeny, Phylogeny, and Historical Development*; Strauss, S., Ed.; Ablex: Norwood, NJ, USA, 1983; pp. 1–27

56. Posner, G.J.; Strike, K.A.; Hewson, P.W.; Gertzog, W.A. Accommodation of a Scientific Conception: Towards a Theory of Conceptual Change. *Sci. Educ.* **1982**, *66*, 211–227. [CrossRef]

57. Strike, K.A.; Posner, G.J. A revisionist theory of conceptual change. In *Philosophy of Science, Cognitive Psychology, and Educational Theory and Practice*; Duschl, R.A., Hamilton, R.J., Eds.; Sunny Press: Newark, NJ, USA, 1992; pp. 147–176.

58. Vosniadou, S. Capturing and modelling the process of conceptual change. *Learn. Instr.* **1994**, *4*, 45–69. [CrossRef]

59. Vosniadou, S.; Skopeliti, I. Conceptual change from the framework theory side of the fence. *Sci. & Educ.* **2014**, *23*, 1427–1445.

60. Vosniadou, S. Initial and scientific understandings and the problem of conceptual change. In *Converging Perspectives on Conceptual Change*; Amin, T.G., Levrini, O., Eds.; Routledge: London, UK, 2017; pp. 17–25.

61. diSessa, A.A.; Sherin, B.L. What Changes in Conceptual Change? *Int. J. Sci. Educ.* **1998**, *20*, 1155–1191. [CrossRef]

62. diSessa, A. A Bird's-Eye View of the "Pieces" vs. "Coherence" Controversy (from the "Pieces" Side of the Fence). In *International Handbook of Research on Conceptual Change*; Vosniadou, S., Ed.; Routledge: New York, NY, USA, 2013; pp. 35–60.
63. diSessa, A. Knowledge in pieces: An evolving framework for understanding knowing and learning. In *Converging Perspectives on Conceptual Change*; Amin, T.G., Levrini, O., Eds.; Routledge: London, UK, 2017; pp. 9–16.
64. Özdemir, G.; Clark, D.B. An overview of conceptual change theories. *Eur. J. Math. Sci. Tech. Educ.* **2007**, *3*, 351–361. [CrossRef]
65. Amin, T.G.; Smith, C.L.; Wiser, M. Student Conceptions and Conceptual Change. Three overlapping phases of research. In *Handbook of Research on Science Education*; Lederman, N.G., Abell, S.K., Eds.; Routledge: London, UK, 2014; Volume 2, pp. 57–81.
66. Amin, T. Representation, concepts, and concept learning. In *Converging Perspectives on Conceptual Change*; Amin, T.G., Levrini, O., Eds.; Routledge: London, UK, 2017; pp. 129–139.
67. Porter, M.F. An algorithm for suffix stripping. *Program* **1980**, *14*, 130–137. [CrossRef]
68. Wootters, W.K. Entanglement of formation of an arbitrary state of two qubits. *Phys. Rev. Lett.* **1998**, *80*, 2245–2248. [CrossRef]
69. Yule, G.U. On the Methods of Measuring Association Between Two Attributes. *J. R Stat. Soc.* **1912**, *75*, 579–652. [CrossRef]
70. Warrens, M.J. On association coefficients for 2x2 tables and properties that do not depend on the marginal distributions. *Psychomterika* **2008**, *73*, 777–789. [CrossRef] [PubMed]
71. Bonett, D.G.; Price, R.M. Statistical Inference for Generalized Yule Coefficients in 2x2 Contingency Tables. *Soc. Meth. Res.* **2007**, *35*, 429–446. [CrossRef]
72. Zhang, Z.; Shan, T.; Chen, G. Random walks on weighted networks. *Phys. Rev. E* **2013**, *87*, 012112. [CrossRef]
73. Majtey, A.; Lamberti, P.W.; Prato, D.P. Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states. *Phys. Rev. A* **2005**, *72*, 052310. [CrossRef]
74. Briët, J.; Harremoës, P. Properties of classical and quantum Jensen-Shannon divergence. *Phys. Rev. A* **2009**, *79*, 052311. [CrossRef]
75. Lamberti, P.W.; Majtey, A.P.; Borras, A.; Casas, M.; Plastino, A. Metric character of the quantum Jensen-Shannon divergence. *Phys. Rev. A* **2008**, *77*, 052311. [CrossRef]
76. Osán, T.M.; Bussandri, D.G.; Lamberti, P.W. Quantum metrics based upon classical Jensen–Shannon divergence. *Physica A* **2022**, *594*, 127001. [CrossRef]
77. Furuichi, S.; Mitroi-Symeonidis, F.-C.; Symeonidis, E. On Some Properties of Tsallis Hypoentropies and Hypodivergences. *Entropy* **2014**, *16*, 5377–5399. [CrossRef]
78. Emmert-Streib, F.; Dehmer, M.; Shi, Y. Fifty years of graph matching, network alignment and network comparison. *Inform. Sci.* **2016**, *346–347*, 180–197.
79. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Elsevier: Amsterdam, The Netherlands, 2012.
80. Newman, M. *Networks*, 2nd ed.; Oxford University Press: Oxford, UK, 2018.
81. Surov, I.A. Natural Code of Subjective Experience. *Biosemiotics* **2022**, *15*, 109–139. [CrossRef]