# Conceptual Analysis in a Computer-Assisted Framework: Mind in Peirce

Davide Pulizzotto, Jean-François Chartier, Francis Lareau, Jean-Guy Meunier, e Louis Chartrand

LANCI, Laboratoire d'ANalyse Cognitive de l'Information, Université du Québec, Montréal
davide.pulizzotto@gmail.com, chartier.jf@gmail.com, francislareau@hotmail.com, meunier.jg@gmail.com, lochartrand@gmail.com

**Abstract.** Conceptual Analysis (CA) is a matter-of-course practice for philosophers and other scholars in the humanities. Exploring one author?s corpus of texts in order to discover the various properties of a concept is a classic example of CA. Recently, a corpus-based computational framework for CA has been emerging in response to the methodological challenges brought about by the massive digitization of texts. In this framework, CA is approached by implementing a computer-assisted text analysis method, within which algorithms are used to support the various cognitive operations involved in CA. In this article, we focus on the retrieval of relevant text segments for analysis. However, this is a complex issue within a computational framework, since the relation between concept and natural language depends on several semantic phenomena, including *synonymy*, *polysemy*, and *contextualmodulation*. The main contribution of this article is methodological because it explores the computational approach to CA. We present three algorithmic methods, which identify relevant text segments while taking into account various semantic phenomena. The results show the potential of computer-assisted CA, thereby highlighting the need to overcome the limitations of these first experiments. An additional contribution of this work takes the form of knowledge transfer from Artificial Intelligence to the Humanities.

**Abstract.** L'analisi concettuale (AC) è una pratica molto diffusa in filosofia e in altri campi delle scienze umane. Esplorare un corpus di testi di un autore per esaminare le proprietà di un concetto è un esempio classico di AC. Recentemente, un quadro computazionale per l'AC basata sui corpora sta emergendo per rispondere alle sfide metodologiche lanciate dalla massiccia digitalizzazione dei testi. In questo quadro, un'AC è vista come un metodo di analisi del testo assistita dal computer, in cui gli algoritmi usati supportano alcune operazioni cognitive dell'AC. In quest'articolo, ci occupiamo del processo d'identificazione dei segmenti di testo che sono pertinenti all'analisi. Tuttavia, questa è una questione complessa in un quadro computazionale, poiché la relazione tra concetto e linguaggio naturale dipende da diversi fenomeni semantici, come la *sinonimia*, la *polisemia* e la *modulazione contestuale*. Il contributo principale di questo lavoro è di tipo metodologico, poiché esplora l'approccio computazionale all'AC. Presentiamo tre catene di trattamento che identificano dei segmenti di testo pertinenti per una AC, tenendo conto di diversi fenomeni semantici.

I risultati rivelano le potenzialità di un'assistenza computazionale all'AC, determinando così la necessità di superare i limiti di queste prime sperimentazioni. Un altro contributo è il trasferimento di conoscenze dall'Intelligenza Artificiale verso le scienze umane.

## 1 Introduction

### 1.1 Conceptual Analysis

If Conceptual Analysis (CA) is a traditional method of inquiry in philosophy [34]; [4], it is also used in many other disciplines. For instance, it is applied in psychiatry [36], in psychology [72], in political science [28], in pedagogy [31], and in many other fields of the humanities and of the social sciences. It is also a common professional practice for lawyers, journalists, physicians, etc.

In the humanities and especially in philosophy, one of the most common forms of CA consists in exploring a corpus of texts by a single or by multiple authors in order to discover the various dimensions or properties of a complex mental process, that is, conceptualization, whose result is called a *concept*. In natural languages, a concept is often conveyed by a *conceptual expression*, which can have various linguistic forms. The study of the conceptual expression of *evolution* and its properties in Darwin's work, or the concept of *beauty* in French novels of the nineteenth century, are examples of CA. Thus, the first step of a CA is to find the text segments that are relevant for the analysis of the concept studied, and this, by finding all the pertinent conceptual expressions. However, it remains difficult to establish a widely recognized and employed CA methodology in the humanities, especially because of the multiple and divergent theories of concepts [38]. Therefore, several methods and approaches exist for conducting a CA [3].

### 1.2 Conceptual Expression and Conceptual Content

What is common to many of these approaches is the close relation between concepts and the meaning of words. In fact, debates regarding theories of concepts have focused on the *lexical concept*, that is, a concept that "correspond[s] to a lexical item in natural languages" ([43], 4). This focus finds its explanation in the assumption that the meaning of a word always conveys a concept and that a concept always has its lexical expression. This phenomenon is here called *the standard lexical form of a concept*, that is, a lexical form regularly used to express a concept. This approach, however, limits the understanding of the problem, as there may exist concepts that do not have such a clear correspondence in natural language. In fact, the relation between concepts and the meaning of words is much more complex ([47], 385) and it cannot be asserted that a concept is expressed in exclusive terms by a single lexical form ([47], 390). Often, concepts go beyond a simple lexical form and they may be expressed by a group of words, by a definition, or by some loose thread that emerges from different text segments. For example, the concept of *marriage* may well be expressed by the word itself, or by the definition "legal union between two people" or by different text segments that

define certain properties of *marriage*, such as *communion of property*. These latter examples give rise to lexical expressions and semantic content that do not require the presence of a sort of *standard lexical form* of the concept of *marriage*. In conclusion, we do not take sides here regarding what the nature of concepts are or whether concepts are independent from or dependent on a lexical expression. The focus of this article is on the *conceptual expressions* (words or sentences) that are carriers of *conceptual content*, which is the set of properties or dimensions of a concept. For simplicity, we will continue to use the term "concept" to refer to the "conceptual content." Ultimately, in our corpus-based approach, the standard lexical form is a starting point from which to study a concept.

### 1.3 Corpus-Based and Computer-Assisted CA

In a corpus-based CA, we first have to find text segments that are relevant for the analysis. This means considering the set of all possible relevant text segments of a corpus. When one studies a large corpus, this becomes a more difficult and complex operation. However, new text analysis tools can be used and adapted to the CA. In recent decades, some disciplines such as Text Mining (TM), Natural Language Processing (NLP), Information Retrieval (IR), and Machine Learning (ML), whose development is linked to Artificial Intelligence (AI) [22]; [59], are inviting certain fields of the humanities into a methodological revolution. Nonetheless, this process is still in a phase of uncertainty, as the transfer of knowledge from subfields of artificial intelligence to the humanities is still in progress.

### 1.4 Objectives of the Article

As we will see in the next section, various semantic phenomena are involved in the definition of the relation between concepts and natural language. In a corpus-based CA, such phenomena must be taken into consideration. In this article, we propose three processing chains that provide assistance in identifying relevant text segments for conceptual analysis, while paying particular attention to the following semantic phenomena: *synonymy*, *polysemy*, and *ellipsis*. We will thus present three experiments and the corresponding results obtained with a philosophical textual corpus: *The Collected Papers* of C. S. Peirce. We will also show how some elements used to identify relevant segments provide more general solutions for CA. The main contribution of the article is the bidirectional transfer of knowledge between AI subfields and the humanities. While philosophy needs tools to assist CA, on the other hand, computer science uncovers original issues through the approaches to text it develops.

The article is divided into four main sections: 1) *Related Work*, mostly pertaining to works focusing on the computer-assisted approach in philosophy and on the corpus-based CA method, 2) *Problem and Theoretical Framework*, in which the problem is presented with respect to the concept vs natural language relation and where the theoretical framework supporting our experiments is briefly summarized, 3) *Method and Experimentation*, where we present our three experiments, 4) *Conclusions*, where we discuss the results and limitations of the experiments.

## 2 Related Work

The use of artificial intelligence techniques in philosophy has been the subject of exploration for several years. Bynum and Moor [10] have integrated these techniques into a general method of text analysis. Lawrence et al. [39] have used topical models to extract arguments in 19th-century philosophical texts. Girju and Moldovan [30] have used ML and NLP tools to answer philosophical questions such as the expression of causal relations in texts. Schwartz et al. [65] have used similar techniques to explore the subjective perception of time. Another category of work applies IR techniques for the creation of a "dynamic ontology" for philosophy [8]. Others have explored NLP techniques to visualize semantic networks for philosophers [1]. Some projects exploring automatic recommendation technology, in order to suggest documents (SalVe2) or providing digital editions of large corpora (*Corpus Thomisticum*), have also been created.[1]

Computer-assisted text analysis has also been applied for CA [45]. The first application of such methods was probably with Allard et al. [2], in which a computer-assisted concept analysis of the Quran was provided. McKinnon [44] studied the concept of *destiny* in Kierkegaard's works. More recently, other researches have explored the concept of *language* in Bergson's work [25]; [15], the concept of *evolution* in Darwin's work [61], the concept of *mind* in Peirce [46], the concept of *management* in the works of philosopher Matsushita [18], and the concepts of *mind* and *body* in early China [67]. The majority of these works used concordance (or KWIC) [51] for such a type of conceptual analysis. A concordance is the set of all the text segments of a corpus where a keyword appears. In CA, the keyword of a concordance is usually the standard lexical form of the concept under study. To our knowledge, except for our previous works [53]; [12], all computer-assisted CAs have been accomplished by retrieving relevant text segments only using the concordance technique. That is to say that CA has been limited to the analysis of the standard lexical form of the studied concept. Thus, it does not go beyond work dealing with the complex relation between concept and natural language.

## 3 Problem and Theoretical Framework

Despite its wide use, and perhaps because of it, CA does not have a unique definition. As it has been introduced, CA methods depend on the nature of the concept. Theories of concepts are categorized in many ways and it is not possible to summarize them here.

### 3.1 Concepts vs Natural Language

When we adopt the linguistic paradigm and the idea that natural language is necessary to express and understand concepts [7]; [16]; [19], language becomes the privileged locus for studying it. The use of linguistic material for philosophical purposes and, more specifically, for CA, remains a widespread practice. However, there is no precise methodology and many elements of the analytical process are not obvious. Bluhm for instance, stressed that:

---

[1] See: `http://www.corpusthomisticum.org/`, `https://salve2.wordpress.com/`.

[I]t is not always obvious which linguistic phenomena are pertinent for the analytical process. Thus, if we take seriously the idea of approaching some philosophical problem through an analysis of ordinary language, we first need to clarify which expressions are to be considered at all. ([5], 8)

Bluhm emphasizes an important aspect of corpus-based CA: What are the expressions of ordinary language that should be considered for analysis? What are the linguistic phenomena to explore? In the same work, Bluhm stressed the fact that philosophers have often relied on their insights about the functioning of language to define the relevant expressions to be analyzed according to a research interest. To prevent any error arising from the researcher's personal knowledge of language, Bluhm proposes the use of linguistic corpora for philosophical analysis, such as the *British National Corpus.*[2]

This type of observation reveals an important element for this article, which is the relation between words and concepts. The nature of the relation between concept and language is the object of various theories of the concept. Among many others [11]; [37] that support a strong relation between concepts and conceptual expressions, we opt here for Murphy's cognitive approach due to its inclusiveness [47]. In his approach, it is impossible to univocally define the relation between these two entities. The idea that every word matches a single concept is not acceptable. In fact, several phenomena can prove the contrary, as do synonymy and polysemy. A word can express many concepts or, conversely, more words can express just one concept. In addition, there are concepts that do not have any precise lexical manifestation ([47], 289).

There are thus different semantic phenomena involved in defining the relation that a word can have with a concept. One of those is *synonymy*. In fact, a concept could be expressed by several lexical forms and all of them can be considered as standard lexical forms. For example, another standard lexical form of *marriage* is also the word "wedding." In this case, we have a *one-to-many* relation, that is, a concept expressed in many ways. In a corpus-based CA, finding the synonyms of a standard lexical form of the concept is a necessary operation. For example, if we analyze the concept of *human* in the writings of Aristotle, we must be able to retrieve its potential synonyms, such as "rational animal." Its opposite phenomenon is *polysemy*[3], where many concepts are related to one lexical item in a *many-to-one* relation. From a semiotic point of view, a standard lexical form is not identifiable with a *signifier* (the sequence of characters), but with the two-faced entity of the *sign* (*signifier-signified*) that best conveys the concept. In a computer-assisted CA, it is therefore important to identify only the occurrences of a word that refer to the concept under study. The classical example is the signifier "bank," which can refer to a financial institution or to the side of a body of water. Another example might be that of the signifier "life," because some uses of this word are not relevant to a CA of the concept of *life*, as when it is used to refer to a "characteristic state or mode of existence" (i.e. "Lifestyle"). A rather extensive category of phenomena, called *contextual modulation* [14], emphasizes the role of *context* in determining the meaning of a word. Cruse ([14], 51) suggests that "the meaning of any word form is in some sense different in every distinct context in which it occurs." According to Murphy, this general process of language emphasizes the importance of a "knowledge approach to concepts,"

---

[2] See: `http://www.natcorp.ox.ac.uk/`

[3] For simplicity, we will just use the concept of polysemy, because it is not easy to clearly distinguish between homology and polysemy ([69], 52).

since the linguistic context of a word and the background knowledge of users are involved in both the definition of a word's meaning and in that of a concept. This is a truism for semiotic studies [70].

In particular, one typical phenomenon of contextual modulation is more important for CA and can be understood through the rhetorical phenomenon of *ellipsis*. In rhetoric, ellipsis is a figure of speech that allows the omission of a word or words within a sentence without its understanding being compromised. [47] exposes an example that recalls this phenomenon. In sentences like (1) "The accountant pounded the stake" and (2) "The accountant pounded the desk," the word "pounded" is elaborated on the basis of the linguistic and pragmatic context, using encyclopedic knowledge. In fact, the understanding of these sentences involves extra-textual information. The words "mallet" or "hammer" are a recall cue for the first sentence and the word "fist" is a better recall cue for the second sentence [68]. This is the *rhizomatic* model of the encyclopedia, as defined by [20], and it reveals, more specifically, how a net of different concepts can be involved in the definition of the meaning of a sentence, even if they do not manifest themselves explicitly in the text. This also happens, but in a different way, in theoretical texts, such as philosophical texts, wherein ellipsis must be conceived as a phenomenon characterized by the omission of the standard lexical form of a concept. Thus, in a CA, if we look at the standard lexical form "nothingness" of the concept of *nothingness*, we must consider that the concept can also emerge in text segments in which the lexical form "nothingness" does not appear, even though it remains underlying.

In a corpus-based CA, all these considerations cannot be overlooked. Answering questions like "What is *nothingness* in Kierkegaard's works?" means analyzing all those text segments that concern the concept of *nothingness*. For a CA, it is crucial *to find the most relevant text segments for the analysis*. Some of these segments can be easily identified if we use the *standard lexical form* of the concept, that is, the words that best convey the concept under study. It is difficult to take issue with the pertinence of this kind of text segment in a CA. However, other types of equally relevant segments are not so easily retrievable, for instance the segments where linguistic phenomena like synonymy, polysemy, or contextual modulation are involved. Some of these phenomena may have simple solutions. For example, synonymy or polysemy can be solved by using a dictionary or by means of some encyclopedic knowledge. Other phenomena, on the other hand, do not have such simple solutions, such as various cases of contextual modulation.

### 3.2   Distributional Hypothesis and Semantic Vector Space Models

If the analysis is based on a small set of texts, this kind of operation does not cause major problems. The identification of relevant segments for CA becomes complex and laborious when the corpus is large. Owing to the advancement of AI research, there are several tools that can be used in text analysis. The computational approach employed in this article is based on a well-defined linguistic theory which derives from structuralism: the *distributional hypothesis*. The distributional paradigm, developed by Zellig Harris and grounded on Leonard Bloomfield's works, is essentially based on the notion that words that occur in the same contexts tend to have similar meanings [32]. Firth coined this principle with the following famous sentence: "You shall know a word by the company it keeps" [27]. In other words, two words tend to be similar if they often stand in the vicinity of the same words.

This theory has paved the way for the *Semantic Vector Space Model* (SVSM) [62], a mathematical model for representing text documents in a vector space. The hypothesis of this model is that the meaning of a text may be represented by means of a vector in a *hyper-dimensional space* whose coordinates derive from the frequency of words the text contains. The simplest SVSM is thus based on the frequency of words in texts and it identifies similar texts by the amount of words they share.

## 4 Method and Experimentation

Each experiment has methodological specificities that will be presented separately. However, there are many shared methodological elements which will be outlined below. Overall, the three experiments present elements that can be used in a more general computer-assisted text analysis methodology.

### 4.1 Corpus and CA

Let us then introduce the most general methodological elements. The corpus under study is the *Collected Papers* (CP), one of the largest collections of C. S. Peirce's writings, containing eight volumes published by the Harvard University Press between 1931 and 1956 [50]. These writings contain approximately 3,000 pages, with 5,163 paragraphs. Being one of the founding fathers of American pragmatism, Peirce still remains an unavoidable reference in philosophy and semiotics.

The concept used for our experimentation is *mind*, which is one of the most important of Peirce's entire work [55]. Studying this concept in Peirce's works is a complex operation, since it is one of the most commonly studied subjects in philosophy [60]. However, getting into the various philosophical theses or debates about the concept of *mind* (the mind-body problem, the relation with consciousness, etc.) is not an objective of this article.

Analyzing the concept of *mind* means to start from a standard lexical form of the concept, that is, the word "mind." However, in order to identify the standard lexical form of the concept, a simple morphological analysis of the word "mind" must be taken into account to disambiguate it. According to the Oxford Dictionary,[4] the word resulting from the chain of characters making up the lexical unit "mind" can be both a noun and a verb. Usually, the uses of the verb "mind" in sentences such as "I do not mind the noise during the day" or "Do you mind if I ask you one more thing?" do not concern the concept of *mind*. Furthermore, the noun "mind" has multiple meanings,[5] such as "recollection or remembrance," "opinion or sentiment," "inclination or desire," etc. Among these, the meaning "the human faculty to which are ascribed thought, feeling, etc." is the most related to the concept of *mind*. This analysis has been taken into account in the experiments.

### 4.2 Pre-Processing

Each experiment we conducted required pre-processing of the textual data. In order to prepare the corpus for computer-assisted CA, it is necessary to perform some operations that allow us to extract the linguistic information needed for the analysis. These operations are the

---

[4] See: https://en.oxforddictionaries.com
[5] https://www.collinsdictionary.com/dictionary/english/mind.

following: 1) *Sentence Boundary Detection*, 2) *Tokenization,* 3) *Part of Speech (POS) Tagging,* 4) *Vectorization.*

Boundary sentence detection is a classic NLP operation, which can be solved using different systems such as the direct encoding of knowledge, a rule-based learning system, static maximum-entropy learning algorithms, etc. [71]. We have used a rule-based learning algorithm[6] and, at the end of this process, we found 44,814 sentences. The tokenization operation consists in identifying each orthographic unit of each sentence. The POS tagging process was performed on the basis of *The Penn Treebank* project [42], which is one of the major references for English. This operation consists in the annotation of each orthographic unit according to its morphological category, such as names, pronouns, verbs, adjectives, numbers, etc. This allows the elimination of some categories of words such as determinants, prepositions, and pronouns, which are not interesting for the analysis. Instead, nouns, verbs, modals, adjectives, adverbs, proper nouns, and foreign words were retained. At the end of this process, we identified 9,668 distinct words, which are called *types*.

Finally, the last operation was the mathematical modeling of each sentence into a vector space. In this model, each sentence is encoded by a vector whose coordinates correspond to the TF-IDF weighting pattern of the words occurring in that sentence. Specifically, this weighting function calculates the normalized frequencies of the words of each sentence [63]. At the end of the process, a matrix $M$ was built containing 44,814 rows corresponding to the sentences found in the corpus and 9,668 dimensions, corresponding to the types.

### 4.3 Other Shared Methodological Elements

To assist the interpretation and evaluation of the results obtained in the three experiments, we relied on a computer-assisted tool and on a human analysis of the concept of *mind* in Peirce's *Collected Papers*. The first one is a classic tool of IR, the *cosine computation* among vectors, which is generally used to select relevant documents. For each experiment, a *query vector* was calculated in order to select some representative sentences. The second was a short qualitative description of the concept of *mind* in Peirce, based on the work of some philosophers who have already studied the topic. This was used as a sort of qualitative benchmark for evaluating the results of each experiment. We provide this description below.

### 4.4 Mind in Peirce

Studies about the concept of *mind* in Peirce's writings are numerous and it is not easy to give an overview of it. However, a brief description of the concept of *mind* was used to evaluate and interpret the segments retrieved in the experiments. In general, we can state that Peirce adopts a real "semiotic approach to the human mind" ([13], 508). For Peirce, "because the totality of the mind's manifestations are signs, we are warranted in identifying mind with semiosis" (CP 5.313). Thus, the study of the mind also concerns the study of the "logic of symbols and signs" ([52], 201). Another important aspect is the relationship between the concept of *mind* and *consciousness*. For Peirce, there seems to be no distinction between the content of consciousness and the manifestation of mind, since both of them are resolved in the sign resulting from an inference (CP 5.313). However, the essence of

---

[6] See the *segtok* tool for python.

*mind* is not *conscience*, but *purpose*, or *final causation* ([66], 554). But what best describes Peirce's idea of mind is the concept of *law of mind* that is "the process by which ideas grow" ([29], 90), a kind of continuous *spread of ideas* that constitutes the mind. This brings us to another important element of Peirce's philosophy, that is, the *immaterial dimension* of the mind. For Peirce, the mind is not in the brain, inasmuch as electricity is not in copper wires ([66], 553). The mind is where thought can be expressed, that is, in "paper or other vehicles for preserving and conveying thoughts" ([66], 553). This highlights another element of his philosophy of mind: It is *thought* that constitutes it, thereby dissociating the mind from a specific material substrate. In fact, thoughts, as mind, are not in the brain: "In my opinion it is much more true that the thoughts of a living writer are in any printed copy of his book than they are in his brain" (CP 7.364). This connection between *mind* and *thought* is supported by his sign theory, since "what makes them thought processes is the sign character of the thoughts" ([66], 554). This makes it possible to express and interpret them. Some other aspects of his philosophy of mind are not easy to summarize, such as the triadic relation between *habit-sign-mind*, leading to a conception of "the human mind [as] an incredibly complex and hierarchically ordered network of habits" ([13], 501), or the relation between mind and *reasoning* and its links with *self-consciousness*, *self-criticism,* and *self-control* ([13], 491).

### 4.5    Experiment No. 1: Synonymy

The first experiment dealt with the semantic phenomenon of synonymy in CA. The aim was to explore how a cosine-based IR algorithm can enable us to retrieve synonyms of "mind" in Pierce's *Collected Papers* and to retrieve relevant text segments for a CA of *mind***.** For this experiment, a simple method consisting of two steps was used: first, building a *similarity matrix,* and second, retrieving relevant text segments with a *query vector*.

**Table 1.** Similarity of "mind"

| Word: Cosine Value |
| --- |
| thought (noun): 0.8237 |
| think (verb): 0.8233 |
| really (adverb): 0.8178 |
| do (verb): 0.8088 |
| nature (noun): 0.8013 |
| thing (noun): 0.7990 |
| seem (verb): 0.7973 |
| come (verb): 0.7968 |
| make (verb): 0.7960 |
| reason (noun): 0.7935 |

The similarity matrix *P* was computed from the *M* matrix, and it represents all the relations of similarity of each word with all of the others. This was obtained by multiplying the *M* matrix with its own transpose and by normalizing the result. So, attributes of each vector of *P* are the scalar products of the distribution of a word for the distribution of every other word. We then got the vector *c* representing the word "mind" according to its relations

of similarity. Subsequently, we could analyze the relation between the vector $c$ and all the other vectors representing each word. The results (Table 1) show that the first nouns closest to mind are "thought", "nature", "thing", "reason". It is also interesting to note that the most similar verb to "mind" is "think". In light of Paragraph 4.4, it is not surprising to note that *thought* is linked to the concept of *mind*.

**Table 2.** Closest Sentences to Vector q

| Paragraph Code | Sentence |
| --- | --- |
| CP 2.228 | "Idea" is here to be understood in a sort of Platonic sense, very familiar in everyday talk; I mean in that sense in which we say that one man catches another man's idea, in which we say that when a man recalls what he was thinking of at some previous time, he recalls the same idea, and in which when a man continues to think anything, say for a tenth of a second, in so far as the thought continues to agree with itself during that time, that is to have a like content, it is the same idea, and is not at each instant of the interval a new idea. |
| CP 7.349 | And an idea can not be thought, except when it is present in the mind. |
| CP 1.444 | In its broader sense, it is the science of the necessary laws of thought, or, still better (thought always taking place by means of signs), it is general semeiotic, treating not merely of truth, but also of the general conditions of signs being signs (which Duns Scotus called grammatica speculativa 1), also of the laws of the evolution of thought, which since it coincides with the study of the necessary conditions of the transmission of meaning by signs from mind to mind, and from one state of mind to another, ought, for the sake of taking advantage of an old association of terms, be called rhetorica speculativa, but which I content myself with inaccurately calling objective logic, because that conveys the correct idea that it is like Hegel's logic. |
| CP 8.13 | But observing that "the external" means simply that which is independent of what phenomenon is immediately present, that is of how we may think or feel; just as "the real" means that which is independent of how we may think or feel about it; it must be granted that there are many objects of true science which are external, because there are many objects of thought which, if they are independent of that thinking whereby they are thought (that is, if they are real), are indisputably independent of all other thoughts and feelings. |
| CP 7.353 | And this causation is necessarily of the nature of a reproduction; because if a thought of a certain kind continues for a certain length of time as it must do to come into consciousness the immediate effect produced by this causality must also be present during the whole time, so that it is a part of that thought. |
| CP 8.329 | The idea of the present instant, which, whether it exists or not, is naturally thought as a point of time in which no thought can take place or any detail be separated, is an idea of Firstness. |

Using the noun "thought" as a synonym, a query vector $q$ was computed with which it was possible to select some typical sentences representing the relationship of similarity between these two words. The vector $q$ is the result of the sum of $c$ and $g$, where $g$ is the vector representing the similarities to the word "thought". Using a cosine computation between query vector $q$ and all vectors in the original matrix $M$, we selected some typical sentences that represent the relationship between "mind" and "thought" within the corpus.

As one can see, these sentences show that for Peirce there exists some similarity of meaning between the nouns "mind" and "thought." This kind of similarity is very important in CA, because it enables us to find some particular properties of the concept under study. For example, in Peirce's *Collected Papers*, these two words share the same relation with "ideas", both as a word and as a concept. As mentioned in Paragraph 4.4, this is a central element of the concept of *law of mind*.

### 4.6  Experiment No. 2: Polysemy

The second experiment dealt with the problem of polysemy in CA. As mentioned above, the standard lexical form of the concept was identified as the noun "mind". However, the latter does not have an exclusive relation with the concept of *mind* and can be used in different contexts with reference to different meanings. Polysemy is one of those linguistic phenomena that interfere with the relation between language and concept, also affecting the process of selecting the text segments to be analyzed during a CA. For this experiment, an unsupervised approach was used. To disambiguate, some tools common in ML were employed, such as *clustering* and *silhouette analysis*. In a text mining framework, clustering algorithms group documents that share similar features, usually word co-occurrence patterns, in order to discover the semantic structures that characterize the documents. In our experiment, we used the *K-means* algorithm ([35], 50), a widely employed algorithm for *word-sense disambiguation* tasks [49]; [54]. The main parameter that needed to be tuned in the K-means algorithm was the *k,* which determines the number of clusters. The *silhouette index* was employed to set this parameter. This index is used as an internal validation measure for clustering [58] because it evaluates the quality of a partition by means of two criteria: the compactness and the separation of clusters. Whenever K-means performs a partition, the silhouette index analyzes the quality of the clusters created. The greater is the similarity within the clusters (compactness) and the greater is the distance between the various clusters (separation), the better is the partition [40].

In our experiment, the silhouette analysis evaluated the first 50 possible partitions of all those text segments containing the noun "mind." The K-means algorithm was applied to all of the text segments where the noun "mind" appears. By finding groups in these text segments, the algorithm identifies different semantic structures reflecting different senses or usages of the chosen keyword. Thus, the algorithm could disambiguate the noun "mind" in Peirce's *Collected Papers*. As plotted in Figure No. 1, according to the silhouette index, one of the best partitions is a K-means clustering into 16 groups.

Some groups of the 16-cluster partition are shown in Table 3 and one can see how the concept of *mind* is related with *theory of signs*, *humanity*, *ontology*, *matter*, *ideas*, *thought*, *time*, *law of mind*, *conscience*, *feelings*, etc. In this experiment, different meanings of the signifier "mind" have been searched. However, cluster analysis suggests that in the corpus analyzed there is no *recurrent* semantic structure deriving from the noun "mind" that *does*

*not* concern the concept of *mind*. In fact, these results disambiguate the *uses* of the noun "mind," thus identifying some dimensions of the concept of *mind* which match the conceptual map described in Paragraph 4.4. So, every sentence in which the noun "mind" is present is potentially relevant for a CA regarding the concept of *mind* in Peirce. For some clusters, we provide a sample of the typical sentence that shows some proprieties of the concept of *mind* in Peirce's work. For each cluster, the most typical sentences have been selected by cosine calculation between the $q_i$ vector, used as a query vector, and each sentence $d_i$ in the corpus, where $q_i$ is the *centroid* of each cluster, which is computed using the K-means algorithm. The limits of this method will be discussed in the Conclusions section.

Table 3: Clustering Results of "mind" Uses

| Cluster Code | Paragraph Code: Sentence |
| --- | --- |
| Clust 0 | 8.262: Moreover, the human mind and the human heart have a filiation to God. |
| | 7.380: It is the human mind that is infinite. |
| | 6.95: The first step of Kant's thought – the first moment of it, if you like that phraseology – is to recognize that all our knowledge is, and forever must be, relative to human experience and to the nature of the human mind. |
| Clust 1 | CP 8.29: It is said that Matter exists without the mind. |
| | CP 4.447: It exists only as an image in the mind. |
| | CP 7.342: So that his knowledge of the thing which exists all the time, exists only by virtue of the fact that when a certain occasion arises a certain idea will come into his mind. |
| Clust 3 | CP 6.101: We know very well that mind, in some sense, acts on matter, and matter on mind: the question is how. |
| | CP 4.611: Now whether this particular way of solving the paradox happens to be the actual way, or not, it suffices to show us that from the supposed fact that mind acts immediately only on mind, and matter immediately only on matter, it by no means follows that mind cannot act on matter, and matter on mind, without any tertium quid. |
| | CP 7.369: Nobody would doubt that that was the true account of the matter, were it not that it is contrary to the law of dynamics that mind should act on matter and contrary to the law of purpose that matter should act on mind. |
| Clust 6 | CP 6.18: A feeling is a state of mind having its own living quality, independent of any other state of mind. |
| | CP 5.288: Consider a state of mind which is a conception. |
| | CP 6.70: Every state of mind, acting under an overruling association, produces another state of mind. |

| | |
|---|---|
| Clust 7 | CP 7.515: Now the generalizing tendency is the great law of mind, the law of association, the law of habit taking.<br>CP 6.277: But it differs essentially from materialism, in that, instead of supposing mind to be governed by blind mechanical law, it supposes the one original law to be the recognized law of mind, the law of association, of which the laws of matter are regarded as mere special results.<br>CP 6.612: In my essay "The Law of Mind" I have so described that law. |
| Clust 8 | CP 7.349: Two ideas exist at different times; consequently what is present to the mind in one is present only at that time, and is absent at the time when the other idea is present.<br>CP 7.350: Now so long as we suppose that what is present to the mind at one time is absolutely distinct from what is present to the mind at another time, our ideas are absolutely individual, and without any similarity.<br>CP 7.348: There is a process which can only take place in a space of time; but an idea is not present to the mind during a space of time – at least not during a space of time in which this idea is replaced by another; for when the moment of its being present is passed, it is no longer in the mind at all. |
| Clust 9 | CP 1.310: The first is that of whatever is in the mind in any mode of consciousness there is necessarily an immediate consciousness and consequently a feeling.<br>CP 7.365: What the psychologists study is mind, not consciousness exclusively.<br>CP 7.366: Consciousness, per se, is nothing else: and consciousness, he maintains, is Mind. |
| Clust 11 | CP 7.349: An idea can contain nothing but what is present to the mind in that idea.<br>CP 7.354: This does not mean that I have always had the idea of prussic acid in my mind, but only that on the proper occasion, on thinking of drinking it, for example, the idea of poison and all the other ideas that that idea would bring up, would arise in my mind.<br>CP 7.392: The clustering of ideas into classes is the simplest form which the association of ideas by the occult nature of ideas, or of the mind, can take. |

| Clust 12 | CP 5.288: Now the logical comprehension of a thought is usually said to consist of the thoughts contained in it; but thoughts are events, acts of the mind. |
| | CP 7.339: That is external to the mind, which is what it is, whatever our thoughts may be on any subject; just as that is real which is what it is, whatever our thoughts may be concerning that particular thing. |
| | CP 8.40: But ordinary people say that not merely the real but all that can possibly enter into the mind of man must be within the thought of God in some sense; so that it must be some particular kind of divine thought which constitutes reality; and that particular kind of thought must be distinguished by a volitional element. |
| Clust 14 | CP 3.361: Supposing, then, the relation of the sign to its object does not lie in a mental association, there must be a direct dual relation of the sign to its object independent of the mind using the sign. |
| | CP 4.536: I have already noted that a Sign has an Object and an Interpretant, the latter being that which the Sign produces in the Quasi-mind that is the Interpreter by determining the latter to a feeling, to an exertion, or to a Sign, which determination is the Interpretant. |
| | CP 8.179: The Sign creates something in the Mind of the Interpreter, which something, in that it has been so created by the sign, has been, in a mediate and relative way, also created by the Object of the Sign, although the Object is essentially other than the Sign. |

As one can see, each cluster of sentences illustrates different semantic structures related with the lexical item "mind." They can be interpreted as "thematic" nuances of the meaning of the noun "mind" and therefore as indicating some properties of the concept of *mind*. For example, cluster No. 3 summarizes the *immaterial dimension* of *mind*, cluster No. 7 focuses on the concept of *law of mind*, cluster No. 11 emphasizes the role of the *spread of ideas* in a theory of mind, cluster No. 12 highlights the relationship between *thought* and *mind*, cluster No. 14 stresses the role of *mind* in a *theory of signs*, etc.

### 4.7   Experiment No. 3: Ellipsis

The last experiment aimed to select relevant elliptical segments for CA. Ellipsis can be seen as a particular *contextual modulation* phenomenon involving conceptual expressions. It describes the semantic process of those text segments where a concept is evoked but without any standard lexical anchorage. In other words, this experiment retrieved text segments where the concept of *mind* is evoked but without its standard lexical form.

In particular, we formalized this task as a positive and unlabeled data (PU) classification problem [6]. This kind of method aims to expand the set of positive data from unlabeled data. The problem of identifying elliptical segments is similar to a PU classification problem because it is possible to interpret the group of text segments containing the noun "mind" as the $P$ dataset (positive dataset), which one wants to expand, and to mark as unlabeled dataset $U$ the rest of the corpus. In other words, we wanted to find the set of elliptical

segments $E$, that is, those text segments that, in a PU classification, would be the *expanded positive data* found among unlabeled data $U$.

As proposed by [48], we used a two-step approach: 1) First, we identified a set of negative data, forming the negative dataset $N$. For this task, the DBSCAN clustering algorithm was used for its capacity to detect outlier data [23], [24], which are observation points that are distant from a closer set of observations; 2) Then, we identified $E$ as a subset of $U$ using the *co-training* algorithm, designed by [6] specifically for the needs related to the PU classification problem. In a nutshell, we first identified the most dissimilar or distant cases from $P$, which were marked as negative data $N$. Then, we trained an algorithm in order to classify the $U$ dataset as positive $P$ or negative $N$.

In the first step, the $U$ dataset was divided into 22 sets, to which we added the $P$ set, forming the $D_i$ set. For each $D_i$ set, the DBSCAN parameters (*Eps* and *MinPts*) were tuned in order to obtain the optimized classifier $S_i$, which aims to find a single large cluster with a number of outliers and which does not exceed 20% of the $D_i$ set size. This rate is an empirical parameter of the method. At the end of the process, we got 8,731 segments forming the $N$ dataset, 1,463 forming the $P$ dataset (18 segments that contain the word "mind" have been marked as outliers), and 34,592 forming the $U$ dataset.

In the second step, we first separated the unlabeled data $U$ into two different datasets, to which we added $N$, forming $Z_1$ and $Z_2$. Then, we applied the co-training algorithm as defined by Luo et al. [41]. This algorithm uses two SVM classifiers, which are first trained on a dataset composed by $P$ as positive data and $Z_i$ as negative data. After this first training phase, these classifiers bootstrap one another by exploring the opposite dataset and by iteratively feeding a second phase of dynamic learning. This process therefore allowed us to explore the $U$ dataset through two classifiers which feed one another by means of two different points of view on data, that is, $Z_1$ and $Z_2$. At the end of the process, we got the $E$ dataset, composed of 2,496 text segments. If in a PU classification, $E$ will be the set of segments that extend $P$, which is generally used for subsequent supervised learning tasks. In our experiment, this set was instead studied as a set containing elliptical segments.

A query vector $q$ was constructed computing the arithmetic mean of the $E$ dataset. Using $q$, we retrieved the following typical elliptical segments:

Table 4: Elliptical segments

| Paragraph Code | Sentence |
| --- | --- |
| CP 6.373 | The idea would, therefore, be found in a pure state only in an immediate consciousness which should make no distinction of any kind, whether between subject and object, or of the parts of the object. |
| CP 7.353 | And this causation is necessarily of the nature of a reproduction; because if a thought of a certain kind continues for a certain length of time as it must do to come into consciousness the immediate effect produced by this causality must also be present during the whole time, so that it is a part of that thought. |
| CP 1.220 | Do I mean that the idea calls new matter into existence? |
| CP 7.345 | For if there be an idea of such a reality, it is the object of that idea of which we are speaking, and which is not independent of thought. |

| CP 8.16 | To make a distinction between the true conception of a thing and the thing itself is, he will say, only to regard one and the same thing from two different points of view; for the immediate object of thought in a true judgment is the reality. |
| --- | --- |
| CP 6.214 | It has, therefore, to do something like supposing a state of things in which that universe did not exist, and consider how it could have arisen. |
| CP 4.55 | The idea which is the matter of the belief is suggested by the idea in those judgments according to some habit of association, and the peculiar character of believing the idea really is so, is derived from the same element in the judgments. |
| CP 7.137 | The formal laws do not depend on any particular state of things, and hence we say we have not derived them from experience; that is to say, any other experience would have furnished the premises for them as well as that which we have experienced; while to discover the material laws we require to have known just such facts as we did. |

The results here have been generated exploring the similarity relations between sentences that contain the word "mind" and sentences that do not contain it. As one can see, this group of sentences shows that Peirce expresses properties about *mind* in different ways. Most of them are to be retained in a conceptual analysis that aims to be as exhaustive as possible. For example, some elliptical sentences focus on *consciousness* (CP 6.373, CP 7.53), some on *idea* (CP 1.220, CP 7.345), some on *thought* (CP 7.345, CP 8.16), some on *habit of association* of ideas (CP 4.55), etc.

## 5   Conclusions

Three experiments have been presented in this article, each of them dealing with a particular semantic phenomenon involved in the process of determining a conceptual expression. Each experiment has detected relevant text segments for a CA of *mind* in Peirce's *Collected Papers,* showing that, to be exhaustive, a corpus-based computer-assisted CA has to deal with synonymy, polysemy, and elliptical segments. Results are encouraging as regards the overcoming of our method's limitations.

In the first experiment, the main problem lied in the difficulty of distinguishing different aspects of the relation of similarity between words. As mentioned above, the distributional hypothesis states that words with similar meanings tend to appear in similar contexts. Many works on synonymy are based on this hypothesis [21] and one of the simplest methods is to analyze the similarities between words through their distribution in the corpus [33]; [40]. This is the approach we followed. However, some have criticized this model because it is incapable of distinguishing the various semantic properties of semantic similarity [9]; [26]. Alternative models have been proposed, which, for example, differentiate between synonymy and antinomy [64]. Overcoming these limitations could improve our algorithm for detecting relevant text segments in a context of synonymy.

In the second experiment, the main problem concerned the difficulty of drawing a distinction between a word's different meanings and its different usages. With regard to NLP

and ML, the phenomenon of polysemy is studied in *word-sense disambiguation* research, which is carried out along three main approaches. A first approach is the *Knowledge-Based* approach, where disambiguation is performed by an algorithm that compares the occurrence of a word with the definitions of a dictionary or ontology, such as the Lesk algorithm. Another one is the *Supervised Corpus-Based* approach, where sense-annotated corpora, such as *Senseval* or *SemCor*, are used for training algorithms to disambiguate word senses. A third one is the *Unsupervised Corpus-Based* approach, where the disambiguation is carried out by a *clustering* algorithm which gathers different uses of a word by only analyzing semantic relationships found in the corpus analyzed. In our second experiment, we followed the unsupervised approach, which does not, in contrast to the other approaches, use external knowledge for disambiguation purposes. Dictionaries and ontologies are employed to recognize different meanings of a word, thus also to identify its uses. To our knowledge, however, there is no corpus or external ontology that is capable of assisting a disambiguation process for a philosophical text.

In the third experiment, the method based itself on the postulate that elliptical segments are semantically similar to those segments containing the noun "mind" (positive dataset *P*). This does not permit to retrieve other kinds of elliptical segments. However, we overcame the limitations of our previous works [53]; [12] where the two unsupervised learning methods depended on some more specific subsets of the positive dataset *P*. Conversely, the semi-supervised approach used in this article allows us to identify elliptical segments in a more global way.

Lastly, this article does not distinguish the logical relations between the various segments found. For example, our method does not allow us to detect definitions, inferences, implications, paraphrases, metalinguistic reformulations, etc. All of these topics are the subject of important theoretical debates pertaining to the analysis of concepts in text corpora. This type of task could be assisted with some tools developed in *argument mining* research [57].

The main contribution of this article is the exploration of the computational approach to CA, developing methods and showing its great potentiality. Another contribution is the transfer of knowledge from AI to the humanities, because these processing chains provide assistance in analytical practices which are widespread in the humanities. Moreover, the article also suggests a new area for text mining research, which is computer-assisted CA, and emphasizes theoretical items that are often neglected, such as the difference between concept and lexeme.

# References

1. Alfano, Mark and Andrew Higgins. "Natural language processing and semantic network visualization for philosophers." In *Methodological advances in experimental philosophy*, ed. Eugen Fischer and Mark Curtis. Bloomsbury, (forthcoming).

2. Allard, Michel, May Elzière, Jean Claude Gardin and Francis Hours. *Analyse conceptuelle du coran sur cartes perforées*. Parigi: Mouton, 1963.

3. Beaney, Michael. "Analysis." ed. Edward Zalta. *The Stanford encyclopedia of philosophy*, 2015.

4. Beaney, Michael. *The analytic turn: analysis in early analytic philosophy and phenomenology*. Routledge, 2010.

5. Bluhm, Roland. "Don't ask, look! Linguistic corpora as a tool for conceptual analysis." In *Was dürfen wir glauben? Was sollen wir tun? Sektionsbeiträge des achten internationalen kongresses der*

*gesellschaft für analytische philosophie e.v.*, ed. Migue Hoeltje, Thomas Spitzley and Wolfgang Spohn, 7–15. DuEPublico, 2013.

6. Blum, Avrim and Tom Mitchell. "Combining labeled and unlabeled data with co-training." In *Proceedings of the eleventh annual conference on computational learning theory*, 92–100. ACM, 1998.

7. Brandom, Robert. *Making it explicit: reasoning, representing, and discursive commitment.* Cambridge, MA: Harvard University Press, 1994.

8. Buckner, Cameron, Mathias Niepert and Colin Allen. "InPhO: the Indiana philosophy ontology."*APA Newsletters-newsletter on philosophy and computers* 7, num. 1 (2007): 26–28.

9. Budanitsky, Alexander and Hirst Graeme. "Evaluating wordnet-based measures of lexical semantic relatedness." *Computational Linguistics* 32, num. 1 (2006): 13–47.

10. Bynum, Terrell Ward and James Moor. *The digital Phoenix: how computers are changing philosophy.* Oxford; Malden, MA: Blackwell Publishers, 1998.

11. Chalmers, David and Frank Jackson. "Conceptual analysis and reductive explanation." *Philosophical review* 110 (2001): 315–61. `https://doi.org/10.1215/00318108-110-3-315`.

12. Chartrand, Louis, Jean-Guy Meunier, Davide Pulizzotto, José López González, Jean-François Chartier, Ngoc Tan Le, Francis Lareau and Julian Trujillo Amaya. "CoFiH: A heuristic for concept discovery in computer-assisted conceptual analysis." In *Proceedings of the 13th International Conference on Statistical Analysis of Textual Data*, ed. Damon Mayaffre, Céline Poudat, Laurent Vanni, Véronique Magri and Peter Follette, 1:85–95, 2016.

13. Colapietro, Vincent M. "Inwardness and autonomy: a neglected aspect of Peirce's approach to mind." *Transactions of the Charles S. Peirce Society* 21, num. 4 (1985): 485–512.

14. Cruse, Alan D. *Lexical semantics.* Cambridge: Cambridge University Press, 1986.

15. Danis, Jean. "L'analyse conceptuelle de textes assistée par ordinateur (LACTAO): Une expérimentation appliquée au concept d'évolution dans l'œuvre d'Henri Bergson." Université du Québec à Montréal, 2012.

16. Davidson, Donald. "Thought and talk." In *Inquiries into truth and interpretation.* Oxford; New York: Oxford University Press; Clarendon Press, 1975.

17. Dekang, Lin. "Automatic retrieval and clustering of similar words." In *Proceedings of the 17th international conference on computational linguistics*, 2: 768–74. Association for computational linguistics, 1998.

18. Ding, Xiaojun. "A text mining approach to studying Matsushita's management thought." In *Proceedings of the fifth international conference on information, process, and knowledge management*, 36–39, 2013.

19. Dummett, Michael A. E. *The seas of language.* Oxford; New York: Oxford University Press; Clarendon Press, 1993.

20. Eco, Umberto. *Semiotica e filosofia del linguaggio.* Torino: Einaudi, 1984.

21. Erk, Katrin. "Vector space models of word meaning and phrase meaning: a survey."*Language and linguistics compass* 6, num. 10 (2012): 635–53. `https://doi.org/10.1002/lnco.362`.

22. Ertel, Wolfgang. *Introduction to artificial intelligence.* Londra: Springer London, 2011.

23. Ester, Martin, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu. "A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise." In *Proceedings of the second international conference on knowledge discovery and data mining*, 226–231. Portland, Oregon: AAAI Press, 1996. `http://dl.acm.org/citation.cfm?id=3001460.3001507`.

24. Ester, Martin, Hans-Peter Kriegel, Jörg Sander and Xiaowei Xu. "Clustering for mining in large spatial databases."*Künstliche intelligenz* 12, num. 1 (1998): 18–24.

25. Estève, Raphaël. "Une approche lexicométrique de la durée bergsonienne."*Actes des journées de la linguistique de corpus* 3 (2008): 247–58.

26. Fabre, Cécile and Alessandro Lenci. "Distributional semantics today."*Traitement automatique des langues, sémantique distributionnelle* 56, num. 2 (2015): 7–20.

27.  Firth, John Rupert. *Papers in linguistics 1934–1951*. Londra: Oxford University Press, 1957.

28.  Forsberg, Tuomas. "Normative power Europe, once again: a conceptual analysis of an ideal type*." *JCMS: Journal of Common Market Studies* 49, num. 6 (2011): 1183–1204. `https://doi.org/10.1111/j.1468-5965.2011.02194.x`.

29.  Gilmore, Richard. "Pragmatism and Islam in Peirce and iqbal: the metaphysics of emergent mind." In *Muhammad Iqbal, 88–111. Essays on the Reconstruction of Modern Muslim Thought.* Edinburgh University Press, 2015.

30.  Girju, Roxana and Dan Moldovan. "Text mining for causal relations." In *FLAIRS 2002 Proceedings*, 360–64, 2002.

31.  Göransson, Kerstin and Claes Nilholm. "Conceptual diversities and empirical shortcomings – a critical analysis of research on inclusive education." *European Journal of Special Nea cura di Education* 29, num. 3 (2014): 265–80. `https://doi.org/10.1080/08856257.2014.933545`.

32.  Harris, Z. "Distributional structure." *Word* 10, num. 23 (1954): 146–62.

33.  Hindle, Donald. "Noun classification from predicate-argument structures." In *Proceedings of the 28th Annual Meeting on Association for Computational Linguistics*, 268–75. Association for computational linguistics, 1990.

34.  Jackson, Frank. *From metaphysics to ethics: a defence of conceptual analysis*. Gloucestershire, UK: Clarendon Press, 1998.

35.  Jain, Anil K. "Data clustering: 50 years beyond k-means." *Pattern recognition letters*, Award winning papers from the 19th International conference on pattern recognition (ICPR), 31, num. 8 (2010): 651–66. `https://doi.org/10.1016/j.patrec.2009.09.011`.

36.  Kendler, Kenneth S. and Michael C. Neale. "Endophenotype: a conceptual analysis." *Molecular psychiatry* 15, num. 8 (2010): 789–97. `https://doi.org/10.1038/mp.2010.8`.

37.  Kipper, Jens. *A two-dimensionalist guide to conceptual analysis*. Berlino; Boston: De Gruyter, 2013. `https://doi.org/10.1515/9783110322705`.

38.  Laurence, Stephen and Eric Margolis. "Concepts and conceptual analysis." *Philosophy and phenomenological research* LXVI, num. 2 (2003): 253–82.

39.  Lawrence, John, Chris Reed, Colin Allen, Simon McAlister, Andrew Ravenscroft and David Bourget. "Mining arguments from 19th century philosophical texts using topic based modelling." In *Proceedings of the first workshop on argumentation mining*, 79–87, 2014.

40.  Liu, Yanchi, Zhongmou Li, Hui Xiong, Xuedong Gao and Junjie Wu. "Understanding of internal clustering validation measures." In *2010 IEEE International conference on data mining*, 911–16. IEEE, D2010. `https://doi.org/10.1109/ICDM.2010.35`.

41.  Luo, Na, Fuyu Yuan e Wanli Zuo. "An integration of cotraining and affinity propagation for PU text classification." In *2009 International conference on computer engineering and technology*, 1:150–54. IEEE, 2009. `https://doi.org/10.1109/ICCET.2009.131`.

42.  Marcus, Mitchell, Mary Ann Marcinkiewicz and Beatrice Santorini. "Building a large annotated corpus of english: the penn treebank." *Computational linguistics* 19, num. 2 (1993): 313–330.

43.  Margolis, Eric and Stephen Laurence. *Concepts: core readings*. Cambridge, Massachusetts: MIT Press, 1999.

44.  McKinnon, Alastair. "The conquest of fate in kierkegaard." *CIRPHO* 1, num. 1 (1973): 45–58.

45.  Meunier, Jean-Guy and Dominique Forest. "Lecture et analyse conceptuelle assistée par ordinateur: premières expériences." In *Annotation automatique et recherche d'informations*, a cura di Jean-Pierre Desclés e Florence Le Priol. Cognition et traitement de l'information. Parigi: Hermes - Lavoisier, 2009.

46.  Meunier, Jean-Guy, Dominic Forest and Ismail Biskri. "Classification and categorization in computer assisted reading and analysis of texts." In *Handbook of categorization in cognitive science*, ed. Claire Lefebvre and Henri Cohen, 955–78. Amsterdam: Elsevier Science Ltd, 2005. `https://doi.org/10.1016/B978-008044612-7/50098-6`.

47.  Murphy, Gregory. *The big book of concepts*. Cambridge Mass.: MIT Press, 2004.

48. Nigam, Kamal, Andrew Kachites McCallum, Sebastian Thrun and Tom Mitchell. "Text classification from labeled and unlabeled documents using EM." *Machine learning* 39, num. 2–3 (2000): 103–34.

49. Pal, Alok Ranjan and Diganta Saha. "Word sense disambiguation: a survey." *International journal of control theory and computer modeling* 5, num. 3 (2015): 1–16.

50. Peirce, Charles Sanders. *The collected papers of Charles Sanders Peirce.* Charles Hartshorne e Paul Weiss, ed. Virginia, U.S.A.: InteLex Corp. Charlottesville, 1994.

51. Pincemin, Bénédicte. "Concordances et concordanciers: de l'art du bon KWAC," 33–42. CALS-CPST, 2007. `https://halshs.archives-ouvertes.fr/halshs-00356008/document`.

52. Powell, Sumner. "Charles S. Peirce, semiosis e the 'mind.'" *ETC: a review of general semantics* 10, num. 3 (1953): 201–8.

53. Pulizzotto, Davide, José A. Lopez, Jean-François Chartier, Jean-Guy Meunier, Louis Chartrand, Francis Lareau and Le Tan Ngoc. "Recherche de «périsegments» dans un contexte d'analyse conceptuelle assistée par ordinateur: le concept d'«esprit» chez Peirce." In *JEP-TALN-RECITAL 2016*, 2:522–31. Parigi: Association Francophone pour la Communication Parlée (AFCP) et Association pour le Traitement Automatique des Langues (ATALA), 2016.

54. Purandare, Amruta and Ted Pedersen. "Word sense discrimination by clustering contexts in vector and similarity spaces." In *Proceedings of the eighth conference on computational natural language learning (CoNLL-2004) at HLT-NAACL 2004*, 2004.

55. Raees, Aisha. "Memory: a semiotic ontology of the self." Doctor of philosophy, Southern Illinois University, 2015.

56. Raees, Aisha. "Peirce's mind." Masters of arts in philosophy, Southern Illinois University, 2007.

57. Reed, Chris. "Proceedings of the third workshop on argument mining (ArgMining2016)." In *Proceedings of the third workshop on argument mining (ArgMining2016)*, 2016.

58. Rousseeuw, Peter J. "Silhouettes: a graphical aid to the interpretation and validation of cluster analysis." *Journal of computational and applied mathematics* 20 (November 1, 1987): 53–65. `https://doi.org/10.1016/0377-0427(87)90125-7`.

59. Russell, Stuart and Peter Norvig. *Artificial intelligence. A Modern approach.* 3rd ed. Pearson, 2016.

60. Ryle, Gilbert. *The concept of mind.* 2009 ed. New York: Routledge, 1949.

61. Sainte-Marie, Maxime, Jean-Guy Meunier, Nicolas Payette e Jean-François Chartier. "Reading Darwin between the lines: a computer-assisted analysis of the concept of evolution in the origin of species," 2010.

62. Salton, Gerard. *Introduction to modern information retrieval.* New York: McGraw-Hill, 1983.

63. Salton, Gerard. *The SMART Retrieval system—experiments in automatic document processing.* Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1971.

64. Scheible, Silke, Sabine Schulte Im Walde and Sylvia Springorum. "Uncovering distributional differences between synonyms and antonyms in a word space model." In *Proceedings of the international joint conference on natural language processing*, 489–497, 2013.

65. Schwartz, Andrew H., Greg Park, Marteen Sap, Evan Weingarten, Johannes Eichstaedt, Margaret Kern, Jonah Berger, Martin Seligman and Lyle Ungar. "Extracting human temporal orientation from Facebook language." In *Proceedings of the The 2015 conference of the North American chapter of the association for computational linguistics-human language technologies*, 2015.

66. Skagestad, Peter. "Peirce's inkstand as an external embodiment of mind." *Transactions of the Charles S. Peirce society* 35, num. 3 (1999): 551–61.

67. Slingerland, Edward, Ryan Nichols, Kristoffer Neilbo and Carson Logan. "The distant reading of religious texts: a 'big data' approach to mind-body concepts in early China." *Journal of the american academy of religion* 85, num. 4 (2017): 985–1016. `https://doi.org/10.1093/jaarel/lfw090`.

68. Sternberg, Robert J. and Peter A. Frensch. *Complex problem solving: principles and mechanisms.* New York; Londra: Psychology Press, 2014.

69. Victorri, Bernard. "La polysémie: un artefact de la linguistique?"*Revue de sémantique et pragmatique*, num. 2 (1997): 41–62.

70. Violi, Patrizia. *Meaning and experience*. Bloomington: Indiana University Press, 2001.

71. Walker, Daniel J., David E. Clements, Maki Darwin and Jan W. Amtrup. "Sentence boundary detection: A comparison of paradigms for improving MT quality." In *Proceedings of MT summit VIII: Santiago de Compostela*, 18–22, 2001.

72. Williams, David M. and Ryan E. Rhodes. "The confounded self-efficacy construct: conceptual analysis and recommendations for future research." *Health psychology review* 10, num. 2 (2016): 113–28. `https://doi.org/10.1080/17437199.2014.941998`.

73. "Full Text PDF." `https://halshs.archives-ouvertes.fr/halshs-00356008/document`.

74. "PubMed Entry." `http://www.ncbi.nlm.nih.gov/pubmed/20142819`.

75. "Snapshot." `https://halshs.archives-ouvertes.fr/halshs-00356008`.

Last consultation URLs: 03/03/2018