Platform for the analysis of written texts in terms of conceptual graphs: study case, Cybernetics of Cybernetics course

José Bermeo
Germán Bravo-Córdoba
Sebastián Contreras
Roberto Zarama
Universidad de Los Andes
Cra. 1 Este 19 A-40
Bogotá - COLOMBIA

{j-bermeo, gbravo, seb-cont, rzarama}@uniandes.edu.co

ABSTRACT

The objective of this work is to observe and analyze the conceptual-relation structures in texts written in Spanish, from the perspective of second-order cybernetics. Texts are shaped by the syntactic structure of the sentences they contain. Conceptual-relations emerge by transforming the text, using a text analysis platform named PAST. The transformation is guided by a set of established rules, permitting to observe a system. PAST can be defined as a powerful operator to transform texts written by the students in a Master’s Program class. The platform is introduced as a tool enabling a graph representation of relationships among the concepts in the text.

The internal procedure of the platform and the graph construction concerns the linguistic analysis. Although the resulting graphic representation does not follow a formal definition, it reflects an iterative construction to define a topology to create a semantic network. This network only makes sense to the author of the text. In other words, the author (observer), based on the cybernetics of the observing systems, uses a tool that enables him to make a second-order observation of a document that he/she has linearly written. The main epistemological concepts mentioned throughout the development of the proposal are discourse and distinction. The contribution of this work is the application of a methodology that conducts a recursive observation of both the direct and indirect relationships established by the words the author puts together in his/her written discourse. In the discussion of results the next phases of this work are presented.

Keywords: Cybernetics, Observer, speech, distinction, semantics, conceptual graph.

1. INTRODUCTION

The objective of a Cybernetics of Cybernetics course (CC) is the questioning, reflection, and construction about the observer’s observing: The interest is that students in the course become able to see the process they follow when they do the observations they do. Such actions are carried out from the perspective of language, questioning the observer who observes the systems, stating his/her own purpose. Regarding the language, the actions of the Observer on first-order cybernetics, ending in second-order cybernetics, are referred to as observer of the observer [6]. The main question that defines the course’s core: Who is observing me as an observer? To answer this question, a methodology was designed, involving games and the construction of a space allowing the interaction among the members of a system (in this case, the CC course), and the theory of graphs [7]. Games are the main pillar of the proposed methodology and it is also the tool used to implement the methodology. Games are understood as the action defining a space where a “world” emerges, bounded by its interpretative context and its own rules.

The CC course defines learning as the acquisition of different concepts and their interconnections, which are corporally absorbed. In the case of CC, the learning and absorbing processes occur as a result of recurrent reading, writing essays, and playing games. The subject of readings is the Observer, his/her emotions and perspectives. In order to analyze the construction of the game space for action and the relationships generated by the methodology, the concept of social networks is used: the design and organization of the players (CC course students) is defined by a regular polyhedron.

During the qualitative revision of the students’ essays, the need to create a text-analysis platform supporting this revision process, while maintaining the author’s (student’s) style and relationship among concepts intact. A first approach used an automated tool to observe and measure structural changes, based on the words and their frequency in the students’ essays [16]. PAST presents then the students’ texts in word-relationships through a graphic representation of the text. The platform transforms the original text into a scheme discovering the relationships among its words. With the word relationships, a graph is depicted as a response file. Thus, we obtain a structure reflecting the relations among the words (nodes), based on linguistic dependencies and in the rules to structure the links (edges). The elements and relations in the graph are interpreted as the distinctive features the author uses to write a text in every single opportunity. Besides of a a different observation of the essay, this analysis also favors the course objective by encouraging the linking of different concepts learned during lectures.

The focus of this document is the text-analysis platform (PAST). To accomplish this objective, some formalization
principles are required: 1. the author (observer) do re-construct his/her written discourse in three opportunities; 2. The platform offers the possibility to build networks of interconnected words. The platform is based on a linguistic context, and a set of rules to identify and relate the networks of words; 3. Rules are defined by the user (professor, analyst and student) and they encourage the generation and extraction of personalized structures. The grouping of such structures constitutes the words network; 4. The suggested transformation to represent the conceptual relations in texts implies the use of graphs as the main data structures in the platform; 5. The main concepts identified are interpreted as the distinctions the author uses in the language of his/her text.

Once formalized these principles, the document is divided in five sections: 1. the epistemological considerations framing the work; 2. the description of PAST as the text analysis platform for the CC course (PAST); 3. the application of PAST and the issued results, including the rules for the analysis, application of the graph and visualization; 4. Discussion of the results; and 5. conclusions.

2. EPISTEMOLOGICAL CONSIDERATIONS

The written text in association with the discourse

According to Foucault, in spoken discourse "The production of discourse is not alienated from the social system in which it is produced" [12]. The CC course defines a social mechanism to regulate the production of discourse and set the boundaries of the space where it is produced. Von Bertalanffy states that the relationship between the language and the vision of the world is not a one-way relationship but a reciprocal one, according to his words "the language structure seems to determine those characteristics of reality to be abstracted and also how thinking categories will be shaped. Thus, the perception of the world determines and shapes the language" [8]. In an adaptation of Bertalanffy’s words, for the observer’s understanding, the discourse (understood as language) expresses the shape of the particular system, subject of the discourse. However, when our discourse is written in the western way: "(...) our language is presented as a causal vision of linear nature" [3]. In other words, it seems that: “we accept the language as an instrument that determines our thoughts and experiences” [2], but an individual’s discourse inevitably reflects the interpretative context of the author as an observer, within the framework of relations of the observer’s social structure. In theory, grammar does not prescribe what we do [10], yet it describes what we say as individuals.

The concept of Distinction applied to the text

Spencer-Brown considers that the starting point of all fields of knowledge requires inventing and making distinctions: "a universe is born when it is separated or alienated from a space and we can draw the limits anywhere we please" [4]. Distinction constitutes the observing act that defines the frontier dividing a space into two sub-spaces, two completely separated continents. We associate the concept of distinction with the concept of ‘reference’ presented in Ogden and Richard’s semantic triangle, which consists in the association among the referent (object), the symbol (in this case the word linked to the object) and the reference between the referent and the symbol. According to the authors, a direct relationship between the referent and the symbol cannot be made without the act of making a reference. In consequence, the described process involves the act of drawing a distinction in the language [Ogden and Richards cited in 13].

In addition, the relationship of the words evidences the individual’s cognitive structure in the language, wherein the act of drawing a distinction in the cognitive level, corresponds to the individual’s concept associations in different levels of abstraction: “For example, the concept “tree” is connected to the more abstract –more general- concept “vegetable”; that is, a tree is a particular case of a vegetable or both. Likewise, the abstract concept “animal” is connected to the more concrete –more specific- “mammal,” and this one to the even more concrete –more specific- “mouse” concept” [17]. Furthermore, a representation of concepts linked by symbols (words of the text) in a non linear way (conceptual graphs), constitutes a cognitive scheme of such abstractions.

Conceptual graphs reflect the distinctions made by each student (observer) each time the submission of an essay is required. In the nodes (unary referent) and arcs topology, distinctions only represent the language associations made by the observer. Therefore, the operations of individual distinction are shaped between networks of observers.

Components of grammar mastering

A number of theories try to explain the grammatical formation of communication through language; in this document, we will refer to “generative grammar”, as proposed by Noam Chomsky. This author states that humankind is homogeneous in terms of linguistic expression, like it is in terms of biology, and that the linguistics of each spoken language is nothing but a combination of basic structures of communication mechanisms. The only genuinely linguistic dominion is the grammatical one, which has three components: a syntactic component, a semantic component, and a phonologic component [15].

Syntax is primordial and it generates the structures (grammatical sentences). Grammatical sentences are made up of individual words and their connections; such distinctions are named constitutive grammar and grammatical dependence respectively. Syntax is built by a basis and a transformational process. The basis is the set of writing rules that generates structure (a group of writing rules that define grammatical relationships and a lexicon that defines the terms and adds semantic and grammatical information). Transformations are the rules that will turn deep structures into superficial structures. The semantic component defines the meaning of those structures, that is to say, the network of words with semantic
relationships, and the phonologic component makes those structures perceptible for interpretation [1]. A hierarchic representation of these grammatical components establishes a dependence hierarchy as follows: syntactic - semantic and phonological component. An additional component, morphologic, has been proposed by L. Boomfield [9].

3. PRESENTATION AND TRANSFORMATION OF THE TEXT ANALYSIS PLATFORM FOR THE CC COURSE (PAST)

PAST is the platform designed to analyze the essays written in the Cybernetics of Cybernetics class 2007. A fraction of the reconstruction of the observer’s experience when observing the CC course is expressed by writing essays. The starting text corresponds to the essays written in the course. Essays are reviewed and returned to students in three different times (t). This action requires two operators. The first operator is iterative and it occurs when: “applying the same operation on the same subject once and again until it is operationally closed and becomes the subject’s own behavior or value” [2]. The second operator, the recurrent one: “[...] shows a system incorporated within another [...] each one with a unique identity, despite the incorporation” [18].

Given the course objectives as the questioning, reflection and construction of the observer’s observer, after the professor’s revision and comments of each of the three essays, the analysis platform will enable the student (observer) to find non-lineal relations in his written discourse. The platform supports and facilitates the evaluation of essays by the professor, and also helps to attain the course’s objective by letting students observe the individual and collective construction of the observer’s observer they are.

The process associated to the text transformation

This process builds a sequence of structures representing the text content in different ways, each one evidencing some of its characteristics. In this section, the transformations obtained in each phase of the process are globally explained:

0. The first step is to eliminate styles, figures, charts, indices, references, etc., leading to a plain text file of the document (1).

1. The first transformation creates a representation of the text maintaining the relationships among words. This transformation only includes linguistic structures existing in the text. By using of a syntactic analyzer called Freeling, a set of linguistic dependence trees are formed, one for each sentence of the text. Each node in the tree contains a word, its associated motto, and its morphologic category, represented by eagle labels; the arcs represent syntactic dependencies, of the “adjective-noun”, “verb-noun” type.

2. The second transformation seeks the full representation of the text and not of its separated sentences, taking into account the relationships among the dependence trees. The transformation of these relationships is based on a set of rules of morphologic composition and transformation (R1), which generate a graph of morphologic structures when applied to each tree.

3. The rules for morphologic composition and transformation (R1) define how the relations of syntactic dependence, hierarchical by nature, are transformed into a graph. These rules define the topology of the graph through two components: the first one is a syntactic dependence structure, represented as a tree, and the second one is its corresponding representation in a graph. The rules match sub-trees in the dependence trees of each sentence of the text, and every time a sub-tree occurs, the words and relationships defined by the rules, are added to the resulting graph.

4. At this point, it is possible to extract words and text relations, in order to make a representation showing how words are related as concepts in the text. For this purpose, the analysis is focused on the nouns, making them the only words that will not be repeated in the graph, as a result rules application. Assuming that words frequency and the way they relate to one another reflects what the author wishes to convey, the nodes and the arcs in the graph are weighted, based on the number of times the rules match the text. The visualization of the resulting graph, based on the weight of the nodes (code of colors) and the arches (thickness), allows the analyst to identify as concepts the nouns that are jointly related in many sentences, since they are represented by nodes with a high weight and having a big number of incident edges.

Thus, the nodes of the morphologic structures graph are the words of the text that occurred in at least one rule of composition. The directed arcs represent the syntactic dependence relationships found in the text, according to the transformations defined by the rules applied during the analysis.

5. The purpose of the third transformation is to focus on the relevance of the relationships among the concepts, making a semantic analysis based on the syntactic structure of the text’s sentences. The suggested process contemplates frontier and relevance analysis of nodes in the graph. The analyst defines two parameters, one for the distance and one for the relevance of nouns. The graph is explored correlating only the nouns, to an extent given by the distance parameter, and filtered preserve only those nouns that have a significant number of relations, defined by the relevance parameter. The distance defines the number of arches between two nouns in the graph; the number of relationships a noun has, defines its relevance in the discourse.

The final result of the distance analysis is a non-directed graph of text’s conceptual proximity approach (4). The nodes represent the concepts in the text. The nodes’ weights correspond to their ordinal organization. The arcs represent the relationship between two concepts and arcs weight represents the structural distance between two concepts. The following figure (1) shows the general scheme of the process:

---

1 Course experience suggests a third iteration as the ideal number of repetitions of an essay. Additional iterations may occur, but it would cost, in terms of time, an occasional extra revision.
A rule is understood as a sub-tree of dependencies, with *eagle-like* labeled nodes. This association with labels in the rule may be generic; it may contain for instance the letter V, which corresponds to a verb, thus matching all verbs. Likewise, a node associated with a VI label matches all verbs in infinitive. To apply the rule, each node of its sub-tree has a numeric ID, in order to define the equivalent structure associated to the rule. The set of defined rules is one of the entries to the association of equivalent structures module of PAST.

Figure 2 shows an example of a rule. This rule contemplates the structural relationship between two subjects (N) and a preposition (S), taking as structural equivalence the relationship N->N. An example is "chess player," which structural equivalence would be a graph conformed by the relationship player->chess. This document shows some of the rules used in the context of CC Course.

The roles associated to the users are those of the supervisor (professor), the analyst and the author of the text himself. While defining the rules, unwanted results may arise due to the empirical nature of the definition process. For the experimental development of the application in the CC course, rules in Table 1 were used; nevertheless, the exercise is susceptible to their partial redefinition.

**Table 1. Dependence and rule**

In the first column of the table 1 represents the dependence tree associated with a rule, and the second column the corresponding representation as a graph. Considering the first row of Table 1, the hierarchical structure of dependence associates a verb with two subjects, whose graph representation shows the relationship between the first subject and the second one.

**A first conceptual graph**

In the grammar realm, it was established that grammatical dependence component reflects the connections used to elaborate a text. As a result of using the syntactic analysis platform, a response file is produced in terms of a graph. The graph structure reflects the grammatical dependencies among words based on the definition of structures. These structures
facilitate the understanding of the implicit semantics in the text, by means of the relationships in the associated dependence trees. The following graph (3) is the result of applying the rules to the dependence tree of the sentence: “the game is the tool for introducing the methodology” (originally in Spanish).

Figure 3. Dependence tree of a sentence and its corresponding graph

In terms of labels, we can observe the identification of the word “game” in the dependence tree, which is "NCMS000", indicating a name (N), common gender (C), no degree assigned (0) and no function assigned (0). This example shows how, in the resulting graph, the word game is coupled with an arc directed to the word tool. This means that the root game is pointed to the dependent tool. This relationship corresponds to the original text that reads: game is the tool, thus representing the syntactic dependence relationships within the text. The graph structure makes an analogue representation of the grammatical dependencies among the words contained in the sentence (the relationships of the dependence tree). Such representations are also visible in the thickness of the arcs; in the previous sentence, the arc joining the word game with the word tool is thicker than the arc joining the word tool and the word incorporate. The number in each node corresponds to the number of matched rules including that word. For instance, the word game appears with a three as number of occurrences, indicating that it matches three of the rules established by the user.

Figure 4a and figure 4b show the results of applying PAST to the paragraph: “In the CC course, we consider learning to be the acquisition and connection of different concepts and that concepts are corporeally interiorized. In the case of the course, the learning and interiorizing processes occur in the recurrent exercise of reading and writing essays related to those readings, as well as designing games and in playing games.”

The distinctions made in the language compose the action dominions. Based on the number of occurrences in the text, the action dominions of previous paragraphs are learning, concepts, interiorizing, readings, essays, and games. At the same time, we can observe how the paragraph’s conceptual graph reflects these action dominions. In the process of transforming a paragraph into a graph, and the later into a conceptual graph, new elements arise that were not visible when the sentence was analyzed. Besides the arcs thickness, their colors represent the weights of the nodes associated with nouns joined in numerous sentences.

In the paragraph, the graph node will be the words: game, design, readings, essays, exercise, interiorizing, learning, acquisition, connection, concepts, process, case, course, CC. These words appeared in the paragraph composition at least once. The syntactic dependence relationships of the paragraph are represented by the arcs according to the rules defined during the analysis.

5. APPLYING PAST IN DIFFERENT MOMENTS

A comparison of the conceptual graphs in three different moments of the course (first, second, and third argumentative essays) enables us to observe the characterization of his/her discourse in the process of wondering who is his/her observer as observer.
As the distinctions in the paragraph of figure 4 emerge, the conceptual graph in time t1 (Figure 5) identifies, amongst others, the distinctions: emotion, observer, individual, dispositions. In addition to the distinctions made by the author, the conceptual graph relates the distinctions found. Generally speaking, at this moment students cannot see clearly the differences between the distinctions they make with the topics of the course and the readings, and concepts and personal appreciations.

The evolution of the exercise in time t2 evidences the appearance of new words and distinctions as well as new relations. In the relations, students recur to the concept of recursion, which: “makes reference to the application of an operation on the result of the application of the operation” [5]. A first recursion for the observer (student) starts to show when the essay in time t1 is taken as historical reference, as compared to that of t2. Besides the distinctions existing in figure 3, in time t2, the distinction “game” stands out.

Similarly, the evolution of the exercise in time t3 (figure 6) evidences the appearance of new words, distinctions and relations; the first idea that arises is how the strengthening of the analysis is proportional to the increasing number of words and their relationships as used within the text, just as is the case in the conceptual map. The differentiation between the most relevant distinctions (the ones with a higher number of relations), and between the concept and the observer’s appreciations justify the relations shown in the conceptual graph. At this moment, the idea of recursion provides insight on how the author’s written discourse reflects the text: “so for me to say that there is recursion, that there is language, in the case of coordination of action, I must be able to make reference to history” [6]. That is, working on the history of the essay analogically represented in a graph, enables the observer’s observer to go back to the action dominions in any given moment, following the process bi-directionally and not in one lineal direction, as it regularly occurs.

![Conceptual graph of Subject X of the CC course in time t1](image1.png)

![Conceptual graph of Subject X from the CC course in time t3](image2.png)

### 6. DISCUSSION OF RESULTS

Although the document, due to space limitations, reflects just one case of the results in terms of the conceptual graph, an elementary analysis of the distinctions reported in all the texts written in the 2007 CC course, found that the number of distinctions follow the chronology of the course. The frequencies and the use of distinctions in the discourse increased, particularly in the transition from second to the third essay.

In this document, we have examined the relationships of the conceptual structures contained in a given text, their distinctions, and the recurrent observations that the observer (student) from the CC course may be able to accomplish on his/her essays. Recurrent observations are facilitated by the syntactic analysis platform, which provides the opportunity to work with a graph as an analogical representation of the text written by the author. Such observations reflect the distinctions of the author’s own “world” and the platform enables the visualization of non-lineal relations in the author’s discourse as they appear in his/her essays. Thus: “our understanding of a universe like that does not derive from discovering its presence, but from recalling what we did to produce it” [4].

Written discourse structures the author’s distinctions in interactions of linear causality: the use of PAST is an invitation to observe bi-directional causalities or causal schemes rather than the conventional one. PAST gives to a text a conceptual organization within the text’s domain, and it shows the different relational structures of the distinctions made by a particular author.

---

2 A conceptual graph enlargement can be seen at appendix A.
Some possibilities for the next phases of the project are: a) Being the purpose of the CC course is to induce students into a self reflection process of observing their own observing process, it will be interesting to let them use PAST as a means to think about their written speech. In other words, if the students are required to produce a text and a metatext on the structure of the former, this metatext can be produced using the results of PAST applied to their texts. This could be used starting from the second version of the essay. b) In a higher recursion level, it could be interesting to encourage a discussion among the students regarding the diverse semantic structures produced individually about a common topic. In this case, PAST could once again be used as a means for guiding these discussions. c) Another possibility is to apply PAST the proposed bibliography of the course. This could also generate particular discussions *vis a vis* about the way the students structure their texts. The structuring way used by the authors could be used as a reference in the course.

7. CONCLUSIONS (TEXT ANALYSIS, RELATION OF PROXIMITY)

Observing the texts linearly does not mean watching the observer’s observer, but observing the observation process. Observing the texts linearly does not mean watching the observer’s observer, but observing the observation process. The rules allows some words (concepts or distinctions the author makes) to connect to other words, creating networks that spread in different directions on a given author’s text.

According to Maturana, for a social system to exist, the recurrence in the interactions resulting from the behavioral coordination of its members must be adapted. Inter-subjective coordination of actions occurs in the same language the subjects live in as observers. Thus, the operations of distinction mold networks of observing-subjects.

The previous exercise shows how, from the perspective of cybernetics, the tool makes possible to confront distinctions. The distinctions made by students (observers) are much more than the sum of their sentences’ meanings and the elements forming their words. In the graph, the observer finds the origin of an idea, providing a new meaning that will be assimilated in the cognitive structure. In other words, a coordinated meaning of his/her discourse expressed in their own words, which will give the observer the opportunity to think of new meanings and make new distinctions in the next iteration when re-writing their previous text.

Conceptual graphs reflect the distinctions made by each student (observer) each time the essay has to be submitted. In node (unitary referents) and arches topology, distinctions only articulate the associations made by the observer, that take place in the language, in a particular moment (t). Two occurrences of the same word can have different meanings, and in a new recurrent operation, these meanings should be communicated within a new network of elements.

Although the analysis of structures is presented for the purpose of understanding the conceptual relations of texts generated during the CC course, the application of the platform is not limited to this course. On the contrary, PAST has proven to be a generic tool, and its operation and results can be applied in a wide variety of fields. Some examples are marketing, consumer’s preferences in terms of experience, or the analysis of the different entries on a patient’s medical history, making them “compatible” and enrich then a health database.

8. REFERENCES

Figure 7. Enlargement of conceptual graph in figure 6.