

The Theoretical Procedures on Linguistics to build an e-service for research results spread

Claudia Wanderley

Researcher at Urban Study Laboratory (Labeurb), Unicamp, BR.

Rua Caio Graco Prado, 70 - Cidade
Universitária "Zeferino Vaz" -Barão
Geraldo - Campinas SP -Brazil

cep 13083 -970

phone/fax: (55) 19-3788-1102/1129

dioutr@hotmail.com

ABSTRACT

The main goal of this paper track is discuss some theoretical issues on Linguistics that come with the Scientific Knowledge Spread on the Web, based on the experience of "Discursive Glossary of the City" project. Our main concern is with Scientific Knowledge Representation, and its integration with language established parameters. Methods to represent this specific knowledge on Web are being studied in Labeurb. As long as reading procedures change, the heuristic approach on knowledge and on language changes as well.

Categories and Subject Descriptors

Heuristic discussions - the role of Philosophy, History, Linguistics, Semantic Web in the production of a knowledge tool for scientists and society.

Text - interpretation, paraphractical procedure, contrast on language models and functions.

Critical un-adaptation of technology - imaginary languages, public policy on knowledge spread, established semantic world.

General Terms

Documentation, Performance, Experimentation, Human Factors, Standardization, Languages, Theory.

Keywords

Discourse Analysis, Semantics, Scientific spread, Heuristics, Web .

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Conference '00, Month 1-2, 2000, City, State.

Copyright 2000 ACM 1-58113-000-0/00/0000...\$5.00.

1. INTRODUCTION

The project "Discursive Glossary of the City", started in January 2000 and it's programmed to be launched on the Web in August 2002. It has the support of Brazilian National Research Center (CNPq), and it's being carried out by a interdisciplinary group of researchers at the Urban Study Laboratory (Labeurb), at the State University of Campinas, Unicamp, Brazil. 1)The information integration, as well as 2) the study of mediation on the Web are the main research topics when we consider the Web as an ideal environment to make scientific findings accessible to other researchers, and to society.

1)The "Discursive Glossary of the City" is a project dedicated to understand basically three layers of scientific findings: A) Knowledge production by researchers; B) Knowledge spread on the Web; C) User's access of knowledge on the web. The main proposal of this project is to build a contact procedure for the user, with RDF data, compatible with the methodology used by researchers at Labeurb, in such a way that even the storing pattern might give cognitive clues about the theoretical core of the glossary. 2) This position statement brings public the theoretical preliminaries/basis of a scientific mediation of the Web, called the "Discursive Glossary of the City" [Gloci], made to enhance the learning of Discourse analysis procedures, with data of a new area of knowledge, called Urban Knowledge and Language, by a specific engineering of data for the user on the Web.

In other words, Gloci is a tool to enhance a reflected relationship between society and research centers on a digital basis, that I understand as a space of *textualization* (textualização in Portuguese), according to Orlandi's[4] text¹ concept.

¹ "In the way I think the Discourse Analysis, when it considers the text as a material form, a concrete manifestation of discourse itself, it makes possible the analysis of its functioning [in Portuguese: *funcionamento*], not by a formal metalanguage use, but by the displacement of the interpretation heuristic place: the analyst does not interpret the text; through an analytical set state, the analyst makes explicit (brings visibility to) interpretation gestures which textualize discursivity, and s/he interprets this analysis results within a theoretical set. As I have insisted, its

2. SOME THEORETICAL ISSUES

"Mais parvenir à articuler ainsi du vrai à propos des matérialités discursives ne va pas sans déplacements de frontières entre les disciplines, affectant profondément en leur régime de vérité, en tant qu'elles y sont provoquées par leurs marges."

M. Pêcheux [8]

The "Discursive Glossary of the City" helps the researcher to have a different picture of the theoretical procedures that s/he is working with. It displaces the relationship among writing, thinking with method, and time. The elements usually thought as conditions of theoretical production become - with the change for a WWW scientific writing - sense effects of a new relationship with Language technologies. Indeed, to understand the birth of a new area of knowledge, the technology that constitutes its writing, and its social accessibility are some of the issues that we discuss theoretically working on Gloci.

To do that, it's important to *un-adaptate*[10] technology and its tools, to bring visibility to other possible ways to deal with language, knowledge, technology, etc. I presented this proposal (of a critical un-adaptation) to enhance a kind of epistemological displacement, of some naturally steady principles to deal with language for instance, which would permit critical approach on the conditions of the production of our ready-to-use tools, and allow us to comprehend new possibilities of interaction with technology, knowledge production, knowledge spread, knowledge data and language. This un-adaptation is being carried out on Gloci project, and the discussion of its problems and results with Semantic Web field is very important, for as researchers we should not work under the evidence of our day-to-day tools, neither under a theoretical straight evidence, but putting in cause their material principles on producing knowledge, the effects of the intellectual work division in a multidisplinar research, and the policies that "naturally" come from it. The un-adaptation of concepts like language, text, interpretation, glossary, etc, and the comprehension of the web environment as a different text materiality lead us to very different way of dealing with scientific data.

2.1 The scientific writing on web

As we consider the possibility to understand Gloci not simply as an electronic publication, but as an interactive publication: not only a reference service, or a tool to access linear reading on Web, but a space to surf which displaces the knowledge writing space boundaries and open Language Studies to re-view the notions of *text*, or language technology *tools*, Language Studies theory is deeply concerned. For me, the procedure of working on a text will strongly work on a digital formulation.

In Discourse Analysis, new technologies come with new text traces, and this traces fusion with what we call "already-there", what is semantically established, or the "old" ways of dealing with

finality it is not to interpret texts, but to comprehend interpretation gestures inscribed into text."(p.78)

texts. So, in this perspective, it is unlikely to consider new technologies as something *per si*, or completely new. As I see, the interpretation of what is a text, and in which materialities can we manage to produce texts, puts language theory in question. And the way to consider the specificity of the changes on digital texts and paper texts is to think on a materialist basis.

In fact, as we brought language theory to understand Gloci, on Web, we worked against the evidence that to do research is to make good questions, and we started to ask the formulation procedure of that question, ask materially about it. It is a radical position, that changes the perspective of our work as Scientists as well as our practice. Pêcheux [8], proposes to "touch language, history and unconscious triple real, without presupposing a more or less general theory of "discourse", what forces us to explore a reservoir of questions that comes around: our encounter problematic territories."

2.2 An interpretation discipline on web

To have an interpretation discipline on web, first of all it is necessary to understand internet environment as a language technology product. This technology is necessarily related with the relationship proposed by Pêcheux [7] among scientific discourse, technical discourse and political discourse, he says that the discourse of the science will become a technical discourse and might work historically as an administration tool.

To think technology, on Discourse Analysis basis, is to think the sense relationship of the scientific discourse and its practices. So, our starting point is the understanding of technological production as a work on language (and about language), as well as an administrative discourse that, at the same time, proposes a linguistic practice and produces a position to a subject.

According to this perspective, the dictionary, the architecture, the painting, the wap, the internet, the psycho-analysis and even the sense of nature - understood as prime material -, for instance, are discourses, within History. These discourses are the material resource that constitutes us - as subjects - in a sense relation with reality, or with what's possible towards reality. We might say that to comprehend a subject in the world is necessary to comprehend how does sense effect of language (and in this specific case: language technologies practices) work.

So, when Discourse Analysis reflects over the effects of scientific knowledge, it's production, it's tools and methods, and the public policy that comes along (and from) all this process on web, what kind of understanding does it bring?

Thinking from an *interpretation discipline* perspective, language technology is a tool made from the study and comprehension of *imaginary-languages*. Orlandi [6] tells us about the imaginary-languages, she says that they're "system-languages, norms, coertions, institution-languages, a-hystorics. Construction. It is systematization that make them loose fluidity and fix themselves as imaginary languages.". Well, new technologies are one of the products of imaginary-languages working on sylicon, energy, web materialities, instead working only on carbon.

These imaginary-languages as a *norm*, in a grammar for instance, circulate in different linguistic theories and models. I will keep the example: grammar might be a book on a shelf, a reference to be researched, a treasure of knowledge about language, a tool for a computer program, a hystorical evidence of national identity, an artefact of linguistic politics, an element

found on cerebral tissue, a cerebral processing pattern unlocalized, a human being intuition, etc. In fact, the study of linguistic normativity works as an interpretation evidence effect in each case, and these interpretations fix themselves and start from this very point, from the singular *materiality* of norm position, which is already in the question of the researcher, and they go forward on a discussion about the nature of "norm" as it was already a real object anchored on the institution that this discursive process produces.

It brings me to consider that the place of the sense *materiality* [5] is a kind of real, from which Science might organize and produce knowledge, analysis, artefacts, and later on might work as social politics [8].

3. GLOCI AND SEMANTIC WEB

"The imaginary language has a return over the real: it shapes it."

Orlandi [6]

For us, it is right where methodology binds these different materialities that we must work toward a comprehension of semantic web discipline. Imaginary language is a product, a material basis to implement a computer procedure, and yet - for a discourse analyst - it's language producing language and they all do return over the real and shape it.

In Gloci project we strongly considered this return to elaborate our knowledge basis. To keep the critical approach on web technology (from a linguistic perspective), the research group decided to invest on a paraphractical procedure and not on a logical oneword-to-oneword, for instance noun-verb-noun, organization. It meant many problems. But this procedure did allow us to think methodologically about the scope of possible ontologies which could be used by a theory that embraces Language, History, Psycho-analysis, or in other words: discourse, ideology and unconscious working in an irregular inter-middle field, on language materiality, towards interpretation practices studies.

3.1 Language and the scope of ontologies

In fact, what makes possible to implement this paraphractical procedure on Gloci project is the notion of text, that we have used as the basic unity to build the structure of Gloci's knowledge base, instead of using *words, concepts, classes*, etc. This choice made a difference on Gloci structure, and in its possibilities, if we consider other linguistic instruments. Gloci is very different from all glossaries, dictionaries and encyclopedias that were thought on a logical basis, or alphabetical basis, which are - for the best we know until now - all of them, and also - as we already pointed out - it differs from a semantic that proposes a closed environment to words and concepts.

Gloci does not deal with a common fund of words/concepts, but with a common fund of texts.

The ontology possible in this case is our discussion at stake, for a possible ontology for Gloci crosses the boundaries of a machine to touch memory, ways to read a file, school process, digital

apartheid, role of a scientific research nucleus in Brazil, Brazilian policies towards science, subjects interpretation possibilities, etc. which are points aroused by the specificity of this language theory tool. For us, as long as there are patterns to work, and a theoretical approach, it is possible to build a kind of ontology. Insofar, a great deal of reflexion is being needed to understand this picture on language studies on web environment, in cause of the specific notion of language within web tools.

So, in Gloci's structure, the main structure to deal with data depends on patterns to interpret, learned in school; depends on the interpretation of a text made by a subject. We assumed that s/he would work searching for a closed and objective meaning, and that researchers should open another interpretation possibility on the text, for this subject, instead of closing it - i.e. paraphractical procedure. For sure, it replaces the relationship with language, and with knowledge - if we think about a question-answer structure, in our project seen as a text-text structure.

In this sense, at first it is hard to envisage how the user would manage to get what s/he wanted, if it is not immediately related with reading practices. But, if we consider that the main theme of the Glossary is "Urban Knowledge and Language", it will be easy to envisage why is so important to put the user in contact with text materialities effects on web, as the paraphractical procedure - which would not be possible to reproduce in a paper publication for example.

As we already said, Gloci differs from other Glossaries structure, because it is dedicated to a specific relationship with language, proposed by Discourse Analysis methodology.

Language has its density, its materiality, and paraphractical procedure does enhance the user to search for an answer experimenting language presence, for 1) the text is taken as the minimum part of the Glossary, 2) the procedure adopted is paraphractical, 3) the theme of the Knowledge base is Urban Knowledge and Language.

At the same time that the web environment permits a project like Gloci in a completely different perspective of traditional Glossaries, surely it brings to Linguistics new questions about language, language processing, language acquisition, and semantics. The discussion of our experience with Semantic Web is one more step towards the comprehension of this difficult relationship among so many disciplines concerning interpretation processes.

3.2 Knowledge spread

Historically, the technological policy strongly invests on a metaphor of an extension of a world semantically normal to a contemporaneous subject. Metaphors as electronic brain, electronic nanny, windows, etc. indicates the need to bring closer to a common sense, to naturalize tools produced with an imaginary language.

On the other hand, and at the same time, another metaphoric movement pushes senses of the subject towards technology, and allows us to talk about carbon subject, biological subject, social subject, or even somethings as silicon subject, optic fiber subject, cyber subject, web subject, etc.

For discourse analysts [11], it's due to public policy that all these metaphores subsist in some theories and/or methodologies concerning language and technology. And it is not really productive to keep these notions within methodology, unless knowledge spread is in question. For us there's no advantage in

"to organism" [*organismar* in Portuguese] technology neither in "to technologize" [*tecnologizar* in Portuguese] the subject, without proper critical approach.

Our specific interest is language materiality. In Gloci project, the interpretation presence lies on the undeniable presence of a meaning that is "already-there". Methodologically, it works as a kind of memory, which will structure the interpretation possibilities and impossibilities. And that's just why paraphrastic procedure matches so well on Gloci project. The effort to reflect about new technologies relation with language turns visible new sense production conditions of technology, the discursive products of the imaginary-language. In Discourse Analysis methodology, it happens at this very same manifold place where language subject and reality are constituted.

4. PRIMER RESULTS AND FINAL CONSIDERATIONS

The result achieved until now is a plastic information integration and a plastic mediation system that allows any theory to become visible in it's own representation on the Web. It might be a root to propose a formal ontology later on, when language processing is capable of dealing with Discourse Analysis methodology principles.

This project is strongly needed in Brazil, for the official scientific spread system comes already with a "method" of representing knowledge, as well as with some "reading" procedures of scientific findings, that most of the time erase the most important object for a Linguist: language itself.

Scientific language has been ambiguously understood as knowing some informations to fill in some forms (or methods to follow), instead of producing knowledge such as Semantic Web, which allows us to comprehend the structure of forms, norms, and therefore re-configure our concepts on information, methodology, science, etc. The development of a scientific language on the Web for researchers, in RDF for instance, it's very important for a Linguist as an heuristic tool that identify the scope of a scientific method - that in fact it is a representation of cognitive procedure (considering its philosophical sense).

In fact, the "Discursive Glossary of the City" project wants to discuss it's actual position towards the possibilities concerning Semantic Web, and some consequences one might learn when the object Language on Web is taken as a serious object. Technological development enhance the discourse analyst to make visible a kind of knowledge production, that wasn't possible to be seen until then. It brings to light relationships within language established in new layers, that ought to be studied. Also, language and methods should be carefully understood by researcher, for they inherit a knowledge in their own structure, and this inheritance without a critic approach might weaken our capacity to deal with new Web challenges.

5. ACKNOWLEDGMENTS

Our thanks to the Brazilian National Research Center (CNPq), the sponsor of this project.

6. REFERENCES

- [1] <http://www.labeurb.unicamp.br>
- [2] Courtine, Jean-Jacques. Définitions d'Orientations Théoriques et Méthodologiques en Analyse de Discours, In *Philosophiques*, vol. IX, n.2, Paris, 1984.
- [3] Foucault, Michel. Nietzsche, Freud e Marx: theatrum filosoficum. Paris 1975. Ed Princípio, São Paulo, BR, 1997
- [4] Orlandi, Eni. Discurso e Texto: formulação e circulação dos sentidos. Ed. Pontes, Campinas, BR, 2001.
- [5] Orlandi, Eni. Análise do Discurso: Princípios e Procedimentos. Ed. Pontes, Campinas BR, 1999. isbn 99-1984
- [6] Orlandi & Souza. A língua imaginária e a língua fluida: dois métodos de trabalho com a linguagem, In *Política Lingüística na América Latina*, Ed. Pontes, Eni Puccinelli Orlandi(org.) Campinas, BR, 1988
- [7] Pêcheux, Michel. Sobre os contextos epistemológicos da Análise do Discurso. In *Escritos 4*, Ed. Labeurb Nudetri - Unicamp, 1999, 27p.
- [8] Pêcheux, Michel. in ouverture du Colloque "Matérialités Discursives"(Nanterre, 24-25-26 Avril 1980,)
- [9] Wanderley, Claudia. O paradoxo da prevenção: a discursividade da AIDS. Master's thesis. Unicamp University, 1999
- [10] Wanderley, Claudia. Scientific Spread and Technology. In I Seminário "Produção e Divulgação do Conhecimento", do Núcleo de Jornalismo Científico (sponsored by Pronex), Nudetri-Unicamp, BR, Jul, 2000.
- [11] Wanderley, Claudia. O corpo, a cidade: repetição. In *Cidade Atravessada: os sentidos públicos no espaço urbano/ Eni Orlandi (org.)*, Ed. Pontes, Campinas BR, 2001

