Epistemology and rhetoric of database-driven translation

Version: 3.3 – February 26, 2009

Konstantin Tovstiadi
University of Oklahoma

> Molecules are omnipresent, but because they are not observable by the naked eye, they must be studied by specialists who are educated in the various ways of revealing their secrets. Likewise, rhetorical structures are present in all forms of text, but they are rarely recognized by the layperson; the rhetorical critic has the specialized training to identify these structures and explain their function.
>
> Leah Ceccarelli[1]

> Étant invisble, il passe pour la nature des choses.
>
> Henri Meschonnic[2]

Abstract

Both translators and translation scholars depend more and more on the digital distribution and storage of texts. Relational databases drive most storage solutions. They are ubiquitous yet invisible to the average user. This paper explores how these databases structure our knowledge and hence have an epistemological significance and how they have a persuasive, or rhetorical, potential. Databases are a symbolic code that favors one possible way of parsing information while suppressing alternative modes of expression. Databases depend on rational logic, completeness, and atomism, and as a result encourage associative arguments, create an illusion of objectivity, and weaken the connection between the text and the author. The goal of the paper is to raise awareness among translation scholars about the power of databases and offer alternative solutions that would be more suitable for storing knowledge rather than information. It also invites other studies of technology as a symbolic code.

The work of a modern translator, and consequently of a scholar studying translation, increasingly relies on computers (Biau and Pym 2006, Bowker 2002, Olohan 2003, Somers 2003). The applications of computers are numerous, ranging from the simplest tasks of using a word processor (in editing a translation) to online collaborative projects using translation memory software. Machines are so ubiquitous that their users are more and more dependent on them. As Boulanger (2005) aptly puts it, "a simple two-hour power shortage in any translation bureau will confirm this fact."

But even if a translation bureau could manage to survive without in-house computers, they would still be dependent on computerized information distribution

networks outside the bureau. The new digital information flow has changed every facet of social actions, and in particular the domains involving texts or other symbolic media. News, photographs, stories, contracts, books, orders, and letters, are almost exclusively processed electronically, at every stage from creating to editing, sharing, publishing, and archiving. The sheer amount of information exchanged requires ways of cataloguing it for easy storage. Not surprisingly, large arrays of information are usually stored in databases. These data banks are governed by strict protocols, rather than saved randomly without a structure. The overwhelming majority of modern data management solution are based on relational databases and run a flavor of the same programming language, SQL (Simple Query Language) (Galindo 2006). These systems are termed *relational* because data in different tables is referenced through a shared field. It is not an exaggeration, then, to say that any professional working with texts inhabits a database-driven world, powered by relational databases and SQL.

This hidden layer of information management is ubiquitous and powerful, yet most users are not familiar with it. In the rare cases that users are aware of it, they don't question relational databases critically, but rather take them for granted as a transparent, innocuous code. This paper investigates how a database-driven world impacts our work as translators and translation scholars. It is based on two assumptions:

1. Relational databases are a mode of organizing information that structures our work (epistemology). They represent only one possible way of organizing information, and as such they cannot be taken for granted.
2. The syntax of database organization is a symbolic code, a structured way of representing ideas, a code that has a persuasive potential (rhetoric). By structuring information in a certain prescribed way and ruling out other ways of organization, they have a rhetorical (persuasive) potential.

The database-driven information flow influences text production, distribution, and eventually translation. As translation scholars, we must question and problematize the epistemology and rhetoric of databases to understand how translation as a practice is changed by them.

To advance my argument, I will begin by describing the logic and syntax of relational databases. I will then analyze them epistemologically (as a way to organize knowledge) and rhetorically (as a persuasive tool). The analysis will allow me to document the ways in which our reliance on databases impacts the way in which we work with texts as translators, translation scholars, and consumers of translations and in doing so, reveal the mechanisms by which this impact may be exercised.

Logic and syntax of relational databases

In a classic textbook on database design, Elmarsi and Navathe (1989) describe a database as "a logically coherent collection of data with some inherent meaning" and note that a "random assortment of data cannot be referred to as a database" (p. 3). In general, any system for storing information, if it is to be accessible and user-friendly, must be based on clearly identified "directives that guide [its] design" (Svetonius 2000: 1).

Relational databases are the most widely used type of information storage, that "underlie almost all the tools we use to organize information… indexing and abstracting services, book sellers, museums and libraries, to name a few, all rely on databases to hold the records of their inventories" (Taylor 2004: 106). The principles governing their structure and use were first put forward by Edgar Codd (1969) in a restricted access

research report for his employer, IBM. It was re-published a year later in a journal article format (Codd 1970), followed by a series of papers through the 1970s (c.f. Codd 1971; see Date 2001, for an exhaustive list and analysis) and finally expanded into a book length volume (Codd 1990).

At several points in the process, Codd explicitly stated the objectives of relational databases. One objective that is present in every list is data independence -- "[a]ctivities of users at terminals and most application programs should remain unaffected when the internal representation of data is changed" (Date 2001: 29). Data independence makes the database itself transparent, invisible; the user is interacting with the interface rather than with the data itself. The intent was to make databases useable for a nonprogrammer, but at the same time that makes the users (including social scientists) oblivious and ignorant of the architecture driving the process.

The structure of the database follows a strict protocol. A database consists of tables, with each row in the table holding a single record. For example, a database driving the library catalog may contain a table storing the names and the dates of birth for the book authors. Each row represents a record for one author; there might be a row containing the information that a certain Karl Marx was born in 1818 and died in 1883. Very importantly, a properly structured database would not store the name as "Karl Marx" and the dates as "1818-1883"; it would present the data in the most atomistic form possible and break it down into the smallest meaningful items. In this case it would use four columns to store the name and the dates: it would have one column for the first name with a record "Karl", one for last name ("Marx"), one for the birth year ("1818") and one for the death year ("1883"). Presenting data in atomistic form is one of the key principles of relational databases. If a query requires a retrieval of the author's last name only, it can easily be recalled from the last name column; if data were not stored in atomistic form and the first and last name were a single string of text, it would have to be parsed on the fly, opening the door for complications and errors.

Another fundamental rule concerns the way different tables relate to each other (hence the word "relational" in the name). Each table contains a column with a unique identifier, such as sequential ID numbers, that can be referenced in other tables. For example, each author in the table described earlier will be given an identification tag – if Marx was entered first into the database, he would receive the ID tag 001. Then in the table describing the books in the library collection, the row containing the information about *Das Kapital* doesn't have to list the full name of the author; it can simply refer to an ID 001. This metonymic structure of relational databases eliminates redundant data, since each bit of information is stored only in one table and is referenced by other tables using an arbitrary code. It also allows for complex data retrieval algorithms to be used when analyzing the data. For example, a search for books by Karl Marx with the word "capital" in the title will search within the book table for titles satisfying this conditions and whose author ID tags refer to the Karl Marx record in the author database. In this case, data from two tables will be pooled to provide the search results; in real life situations, a single search may involve multiple queries pooling data from multiple tables. A lookup of the original edition of *Das Kapital* may draw author's name, title and publication year from one table, the location of the book from another table, and its current check out status from a third.

The introduction of the relational databases has had a tremendous impact on the way information is stored and processed. C. J. Date (2001), one of the most respected

theorists in the field, goes as far as to say that Codd's writings "changed the world as we know it" (p. 1). He also believes that these ideas will continue to be of relevance in the future: "A hundred years from now… database systems will be based on Codd's relational foundation" (p. 1). At the same time he laments that despite the significance of this work, "only a tiny percentage of database professionals have actually read [it]" (p. 3) – historical knowledge is not necessary as long as they are familiar with modern standards and programming languages.

Not surprisingly, then, the computer sources are usually concerned with the technical side of the issue, the status quo of development, such as the explication of the rules described above briefly, as well as database design, management, and the syntax of the language used to interact with databases, SQL (Simple Query Language). Characteristically, the relational rules and SQL are described as a given, without questioning their origin and often not even mentioning the story of how they came about (e.g. Elmasri & Navathe 1989, Brathwaite 1991, Zanasi, Brebbia, & Ebecken 2007).

In library science literature, a similar aphilosophical, ahistorical approach is dominant. Library catalogs are some of the oldest databases in the world, and librarians are still avid users of them; yet databases are treated as an axiomatic foundation on which information systems are built; there is much talk about retrieval techniques, but very little discussion, let alone questioning, of the methods used to store the data (e.g. Stockwell 2001, Rowley & Farrow 2000, Taylor 2004).

Lack of a critical view is characteristic of not only how databases are treated, but apply equally to broader issues, such as the structure of the Internet itself. In another classic textbook, this time for future librarians, Taylor (2004) notes that "the Internet has been likened to a library where all books have been dumped on the floor and there is no catalog" (p. 13). This view is certainly accurate at the macro level; but the chaotic macro structures conceal the strict order of the micro structures. Looking at several news websites together may give the impression of disorder and lack of structure (=chaos at the macro level in between resources), yet each news outlet is organized following a rigid internal structure (=order at the micro level within the resource). The micro structures that rule the storage of data cannot be treated as innocuous, transparent, inconsequential, but must be problematized as limitations and boundaries as to what can be best expressed using them. This paper is an attempt to explore these limitations and to offer alternatives to this mode of managing knowledge.

Not taking these structures for granted means at least three things. First, rather than accepting relational databases and SQL as the de facto standards of information organization, we must admit that these standards are arbitrary and as such they represent just one of many possible ways of data management. Second, since these arbitrary structures are indeed the current dominant standard, they may suppress alternative modes of knowledge structuring. We can safely state that even though we still don't know how information is organized in the brain, we do know, that brains don't run on SQL – in fact, "our brains do not carry language-like labels" at all (Gonthier 2006: 9). Third, these standards represent a symbolic code, and the epistemological and rhetorical potential of this code must be studied. In other words, we must recognize that a code structures our knowledge (epistemology) and through that emphasizes certain modes of expression over others, hence using it is not inconsequential, but persuasive (rhetoric).

Databases and epistemology

The argument that I am advancing in this paper is epistemological, noting that the way of representing knowledge impacts the messages created. It has to do with the potential of a given code to invite certain forms of expression at the expense of other forms.

The significance of this structuring lies not only in the prevalence of one type of argument preferred by a given code but also in the presence of the code itself. The rules of the code are based on a rational logic and as such necessarily bring a degree of rationalization to the discourses created using the code. The more implicit the code is, the more likely it is to be taken for granted, with the consequence that the code's power of structuring thought ignored. We are witnessing a transformation of society where the collective wisdom of the people is moving from books and other printed media to electronic, database-driven media. Just as book publishing has imposed norms and restrictions on the proper way of expressing ideas (Ong 2002), so will the new reality of a server-driven world.

The idea of questioning the code's impact on the message is nothing new. It can be found under various guises in philosophy. It fits within the tradition of the sociology of knowledge where knowledge is not viewed "as a relation between a knower and the world but as a relation between different knowers" (Gonthier 2006: 6). Wittgenstein's (1953/2001) language games, Sapir and Whorf's linguistic relativity (Sapir 1927/1995, 1933/1996; Whorf 1956), Kuhn's paradigms (1960/1996), Foucault's discursive formations and regimes of truth (1970, 1972, 1979) all fit into this tradition. For example, Wittgenstein famously contested the notion that a lion suddenly in possession of speech could talk to humans and give them an insight into the worldview of lions, because learning the human language would have transformed the lion's worldview and the story would not be accurate anymore. The human language is not simply a transparent code; it is a mode of expression that has a fundamental effect on thinking. So too with relational databases: as a widespread mode of expression, they also have a fundamental effect on our thinking.

So what is the epistemology of a relational database? At the most abstract level, databases, like any other cataloging activity, are based on rational thinking, and as such favor rationality over other modes of thinking. They exalt the logic of the text, rather than the pleasure of the text (Barthes 1973).

This rationality is codified by the rules of database design. Two such rules are particularly relevant here: that the data be complete and that the data be presented in the most atomistic form possible.

A user entering records into a database is usually forced to provide full data or the record is not accepted. Form validation, i.e. ensuring that the data entered is complete and in the correct format, is an extremely important aspect of database management (Ullman 2006). For example, if the form is set up to store phone numbers in a hyphenated format (123-456-7890), it should reject user entries without hyphens (1234567890) and make the user re-enter the phone number until it is in the desired format.

Databases work on strict binary logic; either the relationship is present or it isn't. They don't allow for indeterminacy or uncertainty, even though our knowledge is usually incomplete and imprecise:

> … the main methodologies of database design have not paid attention to the modeling of data with uncertainty, although the intent of uncertainty modeling of the real world is rarely absent (Galindo et al 2006: 62).

Data completeness has an obvious advantage – ensuring accuracy of each record and consistency of formatting between different records. But it comes at the expense of leaving out a huge amount of information that doesn't fit the mold.

Besides completeness, relational databases work on the principle of atomism. Data are broken up into the smallest meaningful segments to facilitate data handling. Users don't always have the access to the complete text in the shape it was designed by the author. When a translator works with context-free snippets of texts using a translation memory software tool, the database driven structure of the process is prominent. Thanks to such a fragmentation, there is a current resurgence of equivalence thinking in the localization industry, and utopian hopes of creating multilingual databases that will hold equivalent snippets of texts that will facilitate automated translation (Esselink 2003).

One of the best explanation of why text fragmentation is a problem for translation is found in Meshonnic (1999). Translation studies as a science, translatology (or the French *traductologie*), of which localization approaches are a prime example, focus on the discontinuous in a text -- individual words, unique word meanings, signs and their signifieds, collocations, sentences. This scientism treats translation as a linguistic act, distancing it and even opposing it to literature. Meschonnic's poetics of translation is an alternative to translatology that maintains a close tie with literature rather than with philology and studies the continuous rather than the discontinuous in a given text – rhythm, prosody (p. 23), its semantics (p. 56) and internal coherence (p. 57). A good translation cannot be produced with a focus only on the discontinuous:

> …les critères du bon ou du mauvais ne sont plus des critères simplement philologiques defines par la bonne connaisance de la langue: Amyot et Baudelaire ont fait des fautes, mais leur traduction est bonne. Une traduction sans faute peut être mauvaise (p. 85).[3]

A good translation must take the continuous into account; and hence a translator must use technology that will treat the texts without destroying the continuous. Relational databases are just the opposite -- a technology built with the discontinuous as a cornerstone.

Meschonnic's continuous and discontinuous parallel the dichotomy between knowledge and information. This view can be better explained if we situate these terms in a wider field, and compare the notions of practice and process. The following discussion is based largely on the reworking by Brown and Duguid's (2000) of Lave and Wegner's (1993) discussion of communities of practice. Brown and Duguid make a distinction between two types of social interaction, process and practice. Process is a mechanistic view that sees communication as information transmission. Practice, on the other hand, views interaction as a dynamic learning process, where learning occurs when humans share their experiences and collectively make sense of them. Process is highly structured and follows strict rules, while practice is spontaneous and its flow is dependent on the participants' moment-to-moment decisions. Process is about sharing information, practice is about learning together. Knowledge, then, is information embedded in a social practice. Databases excel at storing information, but are inadequate for storing knowledge.

Not only do databases destroy the fluidity of texts by segmenting them; they also resist modification over time. Once the database programmer determines the initial data structure, it is often too cumbersome to change later on. A new category may emerge that would require a reworking of the old taxonomy, but since most databases need to be always accessible to the users, on-the-fly deep level change is often counterproductive. For

example, one of the oldest classification systems, the Library of Congress call number system, dates back to 1897 and has not undergone deep level change since then. The only way to accommodate new disciplines is to fit them into existing ones as a hierarchical structure, with the original headings as parents and new disciplines as children. As a result, the hierarchy of subclasses in areas with substantial growth (such as sociology, communication, or computer science) is considerably more intricate than in the subclasses with little recent development (such as classic languages).

Epistemologically, relational databases promote rationality, text segmentation, completeness, certainty, and inflexibility. These features can have a positive advantage; but they can be limiting as well. Let us now look at the rhetorical significance of databases. Before describing the rhetorical impact of databases, I need to justify treating them as rhetoric in the first place.

Databases and rhetoric

Communication scholars often assume that everything is communication, even an absence of a message. Translation scholars tend to assume that everything is interpretation, even an individual's internal reasoning. In a similar fashion, rhetorical critics assume that everything has a potential to be persuasive. The title of a famous essay by Richard Weaver, *Language is sermonic* (1990), describes this approach very well: language is not a sterile tool, but always an instrument of argumentation and persuasion.

Databases are a standard, a code for representing information; is it justifiable to look at them as rhetoric? I would like to illustrate my approach by citing a similar study from a branch of rhetorical criticism known as rhetoric of science or rhetoric of inquiry (Bazerman 1987, Ceccarelli,2001, Nelson & Megill 1986, Nelson, Megill, & McCloskey 1987). Scientific discourse has a lot in common with the logic of databases – the same preoccupation with rationality, accuracy, completeness. That is why contrasting the two is not just an idle comparison. The study that I would like to mention deals with a scientific publication manual.

Charles Bazerman (1987) analyzes the seemingly innocuous publication style of the American Psychological Association and demonstrates how the strictly codified structure of the scientific paper prescribed by the manual invites the authors as well as the readers to adopt an attitude of a cataloguer rather than an arguer:

> Within this rhetorical world, the chaos of intellectual difference is eliminated. Individuals accumulate bits, follow rules, check each other out, and add their bits to an encyclopedia of the behavior of the subjects without subjectivity. There is not much room for thinking or venturing here, but much for behaving and adhering to prescriptions (p. 141)

APA formatting style is the standard in most psychological journals, as well as other disciplines like communication and sociology. Bazerman's argument shows how even the simplest code can have a profound impact on discourse. It is not far-fetched, then, to assume that a code as wide-spread as SQL can have a rhetorical significance as well. This is an argument following Ong's (2001) line of reasoning on the "technologizing of the word" -- just as a shift from oral to written is worthy of study, so is the shift from written to digital.

I will argue here that databases function rhetorically in three ways: they encourage associative arguments, create an illusion of objectivity, and weaken the connection between the author and the text. Let me treat these claims one by one.

Databases encourage the creation of lists, structures, and taxonomies, forcing the variety of data into a range of options and imposing relationships that are not necessarily accurate. The focus on relationships brings out associative links between different pieces of data, and through that encourages associative reasoning and associative arguments. Argument from analogy (for example "Patriotism is like loving your family") is not the only possible approach to persuasion. Weaver (1990) has put forth a taxonomy of arguments consisting of three ranks. In this classification, associative arguments it occupies the second rank and is preceded by arguments from definition ("Patriotism is a virtue") and followed only by arguments from authority ("President says patriotism is good"). The quality and ranking of different types of arguments is debatable; but the presence of many types of arguments is obvious. If we are to rely on a technology that favors one type over the others, we at the very least need to be aware of it.

Databases are comprised of complete, checked records; hence the temptation to assume that they are accurate, objective, unbiased records. Here again databases in many ways function the way scientific discourse does. Clerks talking about sales is gossip, a database with transaction sales records is facts. Men talking about engines is idle chat, unconfirmed opinion; an article reporting an experiment testing different motor oils is scientific evidence, truth.

Nelson, Megill, and McCloskey (1987) argue that the dichotomy between truth and opinion is largely artificial and that there is a "need to recognize that rhetoric is reasonable and reason is rhetorical" (p. 33):

> Scholarship uses argument, and argument uses rhetoric. The "rhetoric" is neither mere ornament or manipulation or trickery. It is rhetoric in the ancient sense of persuasive discourse. In matters from mathematical proof to literary criticism, scholars write rhetorically … to compartmentalize truth and opinion, object and subject, substance and form, and the like, rejects the mediation of rhetoric while depending on it (Nelson & Megill 1986: 3, 22).

Flyvbjerg (2001) points out that social scientists suffer from "Cartesian anxiety", wishing to replace the fuzzy context-dependent rhetoric with the certainty of truth. They strive, in Robert Scott's (1967) words, "to smuggle certainty into the realm of human affairs."

Building on the argument that the division between facts and opinion is false, Ceccarelli (2001) reviews the three characteristics of scientific discourse that arguably separate them from other forms of texts. First is the notion of the "recalcitrance of nature" -- the idea I have already discussed at length here, that scientific writing represents an objective representation of reality. Second, she talks about the "exegetical equality" of academic writing -- through the use of a standardized vocabulary and forms of expressions scientists are allegedly able to create context-independent, transparent texts that will be interpreted identically by different readers. Databases can be seen as a recalcitrance of nature, too, since they are collections of "unbiased facts"; they are exegetical because the data in them is complete, fit into standard tables, precise.

Relegating the role of the code to irrelevant background noise is the view that I would like to contest in this paper. In Mantovani's (1996) words,

> …communication environments do not only represent an important background for interpretation of messages which pass through them, but they are themselves meta-messages (p. 69)

Ceccarelli's and Mantovani's ideas can be applied to the new database reality; presenting them as transparent codes that were designed for the unbiased cataloguing of data denies the rhetorical potential of the cataloguing logic.

Having discussed associative arguments and illusory objectivity, let us turn to the last rhetorical feature of databases, the effacement of the author. Pym (2004) suggests that one of the ways in which the new computerized reality impacts translation and translators is the diffusion of origin – a text is altered so many times as it travels from one locale to the next that its source is hardly traceable. Lack of concrete origin and authorship makes it difficult, if not impossible, to apply old source / target dichotomies on the work of modern translators.

The truncation of the text weakens its connection to its origin. Sometimes the reader is not given any information about who the author of the text is at all (Warnick 1998). Lack of a verifiable source may give the users the false impression that the information contained in the database is unbiased, free of individual prejudices imposed by a single author.

In a similar line of argument, when Ceccarelli (2001) talks about peculiar features of scientific texts she notes that they are institutionally driven – the role of the author is less important, authors act not as individual agents but rather as cogs in the machine of science.

Warnick's (1998) "indeterminacy of authorship" (p. 74), Pym's (2001) moving texts without a clearly identifiable source, Ceccarelli's (2001) institutionalization are all facets of the same phenomenon, the effacement of the author. Human writing is an assemblage of ideas from different sources that can't be viewed as original, authored by a single invidividual (Barthes' (1977) death of the author); database-driven texts are often assembled on the fly as well, pieced together from different locations based, custom made based on the user's unique requests.

Dealing with a database-driven world

A patient comes to a doctor, moves his arm up and down, and says, "Doctor, when I do this, it hurts", to which the doctor immediately replies, "Well, don't do it". It would be naïve to say that the best way to deal with databases is to stop using them. Instead, I offer two alternative solutions.

This paper is the embodiment of the first solution -- raising awareness, revealing a code that is ubiquitous yet implicit, a code whose power derives from its very implicitness. It has followed the ethic of a rhetorical critic, who, rather than condemning or exalting discursive practices, reveals implicit means of persuasion in discourse. Such an analysis gives the audiences a more accurate picture of the way symbolic processes affect their lives.

The second solution is about technology that can be added to the relational database standard to move away from its rationality, atomism, completeness; and in doing so, to allow the treatment of the continuous, not just the discontinuous; the storage of knowledge, not just information; and the fostering of a full range of arguments, not just associative arguments. Four related technologies come to mind here: knowledge management systems, fuzzy logic, emergent relationships, and adaptation. Again, let us deal with them one by one.

First, if we view knowledge as information embedded in a practice, then we need a way to digitally store and process knowledge rather than information. This is the turf of knowledge management (Brachman & Levesque 2004, Stefik 1995, Van Harmelen et al

2008), a branch of artificial intelligence research. While relational databases are concerned with the syntax of the relationships between different pieces of data, knowledge management is about the complex semantics of these relationships. Databases are mathematical abstractions of reality, and knowledge systems are semantic approximations of reality (or at least they attempt to be). These systems are still relational in the sense that different pieces of data are referenced to each other; but these links are multidimensional rather than unidimensional. The links resembles the multitude and the parallelism of connections between neurons in the brain, not the drawing of a genealogical tree.

Besides allowing multidimensional links between data, we can also modify the way the data themselves are stored. This is the second technology that can work against the determinism of databases. Databases work on strict binary logic; either the relationship is present or it isn't. They don't allow for indeterminacy or uncertainty, even though our knowledge is usually incomplete and imprecise (Galindo et al 2006: 62).

To allow for the more realistic approximate reasoning rather than the exact reasoning, the formal logic of the databases can be replaced with the *fuzzy logic* of knowledge systems, where "exact reasoning is considered as a specific case of approximate reasoning" and "all problems are problems of degree" (Galindo et al 2006: 5).

A database driven by formal logic may, for example, allow for three values for the variable of age – "young", "middle-aged" and "old", with clearly defined boundaries and no overlap between the categories. In fuzzy logic, the categories would overlap and each data point would be assigned an additional ratio to estimate the certainty of the information that is stored. These ratios help deal with various types of imprecise information – uncertainty, imprecision, vagueness, inconsistency, and ambiguity (Motro 1995). Using fuzzy logic, information storage can more accurately represent our knowledge by incorporating data about the certainty of the information. For example, a doctor looking for records of "young" patients in a medical database will also be given all boundary cases of "middle-aged" that are close enough to "young"; a search for "middle-aged" will retrieve boundary cases of "young" and "old"; thus the imprecision of the query will match the fuzziness of the output.

An alternative way to account for the imprecision of our data is to still use simple relational logic for data storage, but then incorporate complex algorithms at the stage of data analysis. For example, a large online retailer will store information about millions of customers and their transactions, and then use it to give meaningful recommendations for future purchases based on the customer's purchasing history and a comparison with similar customers. These recommendations are imprecise, fuzzy knowledge; there is no guarantee that the customer will like them, but it is worth noting that in this case the uncertainty emerges from a system where at the data storage level there is no uncertainty.

In the three cases described so far, the rigidity of the micro structure (=relational databases used for data storage) is counterbalanced by the flexibility of the macro structures (=knowledge systems, fuzzy logic and emergence). Fourth and last, the synchronic rigidity of databases can also be compensated by a longitudinal fluidity, by adaptation – the data change and evolve according to the interactions with the users. A data source that fails to adapt to the change (perhaps because the database structure doesn't reflect the new reality) falls into oblivion and a newer more capable resource takes its place.

The analysis presented in this paper can be seen as part of a larger project that would problematize the invisible symbolic layers that drive our more and more computerized world, a project that views technology as a code. Looking at the database level is just one of many possible angles. A related question is about artificial languages, in our case programming languages. Most, if not all, key programming languages use English as their foundation. The most widely adopted operating systems and applications are driven by the code written in these languages, and relying on English as a foundation. The implications of these phenomena need to be questioned from a standpoint of a rhetorical critic.

Acknowledgements

I would like to thank Lisa Foster, Laura Gibbs, Gregory Miller, and Lee Williams for their invaluable feedback on earlier versions of this paper.

Notes

1. Ceccarelli 2001: 324.
2. Meschonnic 1999: 98. "Being invisible, it passes for the nature of things."
3. "The criteria of good and bad are not simply philological criteria, defined by the good knowledge of the language: Amyot and Baudelaire have made mistakes but their translations are good. A translation without mistakes can be bad."

References

Barthes, Roland. 1973. *Le plaisir du texte* [The pleasure of the text]. Paris: Editions du Seuil.

Barthes, Roland. 1977. Image, music, text (translated by Stephen Heath). New York: Hill and Wang.

Bazerman, Charles. 1987. Codifying the social scientific style: The APA *Publication Manual* as a behaviorist rhetoric. In John S. Nelson, Allan Megill, Donald N. McCloskey (Eds.), *The rhetoric of the human sciences: Language and argument in scholarship and public affairs* (pp. 125-144). Madison, WI: University of Wisconsin Press.

Biau, Jose Ramon, & Pym, Anthony. 2006. Technology and translation: A pedagogical overview. Full text online at: http://www.tinet.org/~apym/on-line/training.html

Boulanger, Pier-Pascale. 2005. Transdisciplinary translators? Paper presented at the International Conference on Translation and Interpretation, Monterey Institute of International Studies, September 9, 2005.

Bowker, Lynne. 2002. *Computer-aided translation technology: A practical introduction*. Ottawa: University of Ottawa Press.

Brachman, Ronald. J., & Levesque, Hector J. 2004. *Knowledge representation and reasoning*. Amsterdam: Morgan Kaufmann.

Brathwaite, Kenmore S. 1991. Relational theory: Concepts and applications. San Diego, CA: Academic Press.

Brown, John S., & Duguid, P. 2000. *The social life of information.* Boston, MA: Harvard Business School.

Ceccarelli, Leah. 2001. Rhetorical criticism and the rhetoric of science. *Western Journal of Communication, 65* (3), 314-329.

Codd, Edgar F. 1969. Derivability, redundancy and consistency of relations stored in large data banks. IBM research report RJ599.

Codd, Edgar F. 1970. A relational model of data for large shared data banks. *Communications of the ACM, 13*, 377-387.

Codd, Edgar F. 1971. A data base sublanguage founded on the relational calculus. IBM research report RJ893.

Codd, Edgar F. 1990. The relational model for database management: Version 2. Reading, MA: Addison-Wesley.

Date, C. J. 2001. *The database relational model: A retrospective review and analysis*. Reading, MA: Addison-Wesley.

Elmasri, Ramez, & Navathe, Shamkant B. 1989. Fundamentals of database systems. Redwood City, CA: Benjamin / Cummings.

Esselink, Bert. 2003. Localisation and translation. In Harold Somers (Ed.), *Computers and translation: A translator's guide* (pp. 67-86). Amsterdam: John Benjamins.

Flyvbjerg, Bent. 2001. *Making social science matter: Why social inquiry fails and how it can succeed again.* Cambridge: Cambridge University Press.

Foucault, Michel. 1970. *The order of things: An archaeology of human sciences.* New York: Pantheon.

Foucault, Michel. 1972. *The archaeology of knowledge and the discourse on language*. (Trans. A. M. Sheridan Smith). New York: Pantheon.

Foucault, Michel. 1979. *Discipline and punish: The birth of the prison.* (Trans. A. Sheridan). New York: Vintage/Random House.

Galindo, Jose, Urrutia, Angelica, & Piattini, Mario. 2006. *Fuzzy databases: Modeling, design and implementation.* Hershey, PA: Idea Group.

Gonthier, Nathalie. 2006. Introduction to evolutionary epistemology, language and culture. In Natalie Gonthier, Jean Paul van Bendegem, and Diederik Aerts (Eds.), *Evolutionary epistemology, language and culture: A non-adaptionist, systems theoretical approach* (pp. 1-32). Dordrecht, The Netherlands: Springer.

Kuhn, Thomas S. 1960/1996. *The structure of scientific revolutions* (3rd ed.). Chicago: University of Chicago Press.

Lave, J., & Wenger, E. 1993. *Situated learning: Legitimate peripheral participation.* New York: Cambridge University Press.

Mantovani, Giuseppi. 1996. *New communication environments: From everyday to virtual.* London: Taylor and Francis.

Meschonnic, Henri. 1999. Poétique du traduire [Poetics of translating]. Paris: Verdier.

Motro. A. (1995). Management of uncertainty in database systems. In W. Kim (Ed.), *Modern database system: The object model, interoperability and beyond.* Addison-Wesley.

Nelson, John S., & Megill, Allan. 1986. Rhetoric of inquiry: Projects and prospects. *Quarterly Journal of Speech, 72,* 20-37.

Nelson, John S., Megill, Allan, & McCloskey, Donald N. 1987. Rhetoric of inquiry. In John S. Nelson, Allan Megill, Donald N. McCloskey (Eds.), *The rhetoric of the human sciences: Language and argument in scholarship and public affairs* (pp. 3-18). Madison, WI: University of Wisconsin Press.

Olahan, Maeve. 2004. *Introducing Corpora in Translation Studies*, London and New York: Routledge.

Ong, Walter.2002. *Orality and literacy: The technologizing of the word.* (2nd ed.). New York: Routledge.

Pym, Anthony. 2004. *The moving text: Localization, translation, and distribution.* Amsterdam: John Benjamins.

Rowley, Jennifer, & Farrow, John. 2000. *Organizing knowledge: An introduction to managing access to information* (3rd ed.). Aldershot, England: Gower.

Sapir, Edward. 1927/1995. The unconscious patterning of behavior in society. In Ben G. Blount, (Ed.), *Language, culture, and society: A book of readings (2nd ed.).* (pp. 29-42). Prospect Heights, IL: Waveland Press. (Reprinted from *The unconscious: A symposium*, by E. S. Drummer, Ed., 1927, New York: Knopf).

Scott, Robert L. 1967. On Viewing Rhetoric as Epistemic*. Central States Speech Journal, 18*, 9-17.

Sapir, Edward. 1933/1996. Language. In Ben G. Blount, (Ed.), *Language, culture, and society: A book of readings (2nd ed.).* (pp. 43-63). Prospect Heights, IL: Waveland Press. (Reprinted from the *Encyclopedia of the social sciences,* pp. 155-69, by Edwin A. Seligman, Ed., 1933, Macmillan Publishing Company).

Somers, Harold (Ed.). 2003. *Computers and translators: A translator's guide* (Benjamins Translation Library vol. 35). Amsterdam: John Benjamins.

Stefik, Mark. 1995. *Introduction to knowledge systems.* San Francisco: Morgan Kaufmann.

Stockwell, Foster. 2001. *A history of information storage and retrieval*. Jefferson, NC: McFarland.

Suh, Nam P. 2005. *Complexity: Theory and applications.* New York: Oxford University Press.

Svenonius, Elaine. 2000. *The intellectual foundation of information organization.* Cambridge, MA: MIT Press.

Taylor, Arlene G. 2004. *The organization of information* (2nd ed.). Westport, CT: Libraries Unlimited.

Ullman, Larry. 2006. *MySQL*. (2nd ed.). Berkeley, CA: Peachpit Press.

Van Harmelen, Frank, Lifschitz, Vladimir, & Porter, Bruce. 2008. *Handbook of knowledge representation.* Amsterdam: Elsevier.

Warnick, Barbara. 1998. Rhetorical criticism of public discourse on the Internet: Theoretical implications. *Rhetoric Society Quarterly, 28*, 73-84.

Weaver, Richard. 1990. Language is sermonic. In P. Bizzell & B. Herzberg (Eds.), *The rhetorical tradition: Readings from the classical times to the present* (pp. 1044-1054). Boston: St. Martin's Press.

Whorf, Benjamin Lee. 1956. *Language, thought, and reality.* Cambridge, MA: The M.I.T. Press.

Wittgenstein, Ludwig. 1953/2001. *Philosophische Untersuchungen / Philosophical Investigations*: The German text, with a revised English translation (3rd ed., translated by G. E. M. Anscombe). Oxford: Blackwell.

Zanasi, A., Brebbia, C.A., & Ebecken N.F.F. (Eds.). 2007. *Data mining VIII: Data, text and web mining and their business applications.* Southampton, UK: WIT Press.