



Writing and monitoring in International Standardization, Theoretical Choices and Methodological tools

Brigitte Juanals, Jean-Luc Minel

► To cite this version:

Brigitte Juanals, Jean-Luc Minel. Writing and monitoring in International Standardization, Theoretical Choices and Methodological tools. International Institute of Informatics and Systems. IMETI 2011, Jul 2011, Orlando, United States. 1, pp.82-87, 2011. <halshs-00610032>

HAL Id: halshs-00610032

<https://halshs.archives-ouvertes.fr/halshs-00610032>

Submitted on 20 Jul 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Writing and Monitoring in International Standardization, Theoretical Choices and Methodological Tools

Brigitte Juanals

MoDyCo (UMR 7114) – Université Paris Ouest Nanterre La Défense – CNRS
Nanterre, 92001, France

and

Jean-Luc Minel

MoDyCo (UMR 7114) – Université Paris Ouest Nanterre La Défense – CNRS
Nanterre, 92001, France

ABSTRACT

Many organizations are in charge of global security management. This paper outlines and argues for the construction of a theoretical and methodological framework in order to critically assess the new technopolitics currently being developed in the field of global security and which are materialized in standards. The main purpose is to design both a methodology and specific text mining tools to investigate these standards. These tools will be implemented in a platform designed to provide cartographic representations of standards and to assist the navigation of an end-user through a corpus of standards.

Keywords. Standardization, text mining, ontologies

1. INTRODUCTION AND CONTEXT OF THE RESEARCH

The research developed in this paper relates to the industrial standardization of security and risks, located at the intersection of global security and international standardization. The technical and political aspects of these two areas give rise to technopolitics, which we contend will play a leading role in the field of industrial regulation in the XXIst century, causing major industrial, economic and geopolitical impacts.

Many organizations are in charge of global security management. To address this issue, ISO (International Standards Organization), the main international organization for technical standardization, has launched a set of studies in the wake of the ANSI program “Homeland Security Standard Panel” (law of 2002, November 25), subsequently adopted by European and national standardization organizations.

This paper outlines and argues for the construction of a theoretical and methodological framework in order to critically assess the new technopolitics currently being developed in the field of global security and which are materialized in standards. The main purpose is to design both a methodology and specific text mining tools to investigate standards. These tools are designed as heuristics which enable reformulations, semiotic transfer (texts to diagrams), and comparison between texts.

This work is part of a scientific research project called “NOTSEG”¹ which stands for “Standardization and Global

Security, The formulation in Standardization of the global security concept”. The partnership includes academic and industrial partners: two research centres, MoDyCo from the University of Paris Ouest Nanterre La Défense - CNRS, CQP2I from the Technological University of Compiègne, AFNOR, the French national standardization body, and the company Sector SA, specialized in decision-making and studies in the area of technological and organisational risks.

The NOTSEG project mission is to draw up the cartography of existing normalization frameworks in the field of security and crisis management. The project comprises several successive stages, namely: i) inventorying the list of standards to be studied; ii) analysing existing standardization frameworks; iii) studying and selecting variables; iv) analysing and monitoring a corpus of standards using text mining tools. The methodology of these tools is the topic of this paper.

Recent evolutions in standards are the core of our analysis. The processes of international industrial standardization apply not only to the artefacts of technical devices (in technical standards), but also, since the year 2000, to organizational methods and the evaluation process of these devices, including state regulations, especially in standards of management. Domains such as “business continuity” and “resilience” now possess standards of ‘security management’.

The rest of this paper is organised as follows. Section 2 explains the theoretical approach and issues. Section 3 details our hypothesis and methodological approach. Section 4 describes the experimentation in progress, section 5 details the platform specifications and finally section 6 presents the conclusion.

2. THEORETICAL APPROACH AND ISSUES

Texts on industrial standards are here considered as means of validating and communicating technical choices, knowledge, and professional practices, as manifested in their cultural and industrial contexts within the wider international and economic situation. We aim to identify, in a specific sector of activity, the conditions and mechanisms that are conducive to new standards contributing to focusing and communicating certain practices, processes, modes of organization and socio-technical arrangements, on the European or international level. In the second step, it will be necessary to investigate whether these new standards really play a role in the evolution of the above-mentioned elements. If this is the case, changes in stakeholders’ knowledge and professional practices will then be considered.

¹ The NOTSEG Project (www.notseg.fr) is funded by the French National Research Agency (ANR) over a 3-year period (2009-2012).

In international standardization, the process of establishing each standard is dependent on the conditions of production goals and modalities of use anticipated by the standard. The development of standards implies the analysis, case by case, of the means of action and coordination of activities by organizations and communities [1].

In what way do standards contribute to the communication of knowledge in a specific sector of activity? Standards encode knowledge, modes of organization (such as the «Plan-Do-Check-Act model», a widely used corporate model that structures the management process), approaches (to the nature and evaluation of risks, for instance), or procedures that need rethinking.

In the writing process of standards, our aim is to investigate, beyond the institutional display of a “consensual” operation, a pragmatic dimension in a given sector of activity (its technical committee), the goals which govern their construction, the categories of stakeholders involved, the motivations which influence choices (in terms of writing the contents), the participants in workgroups, and the adjustments that take place between stakeholders from different cultures and languages. We wish to point out that these stakeholders have professional practices in the same domain which can be similar, different, complementary but seldom conflicting.

What conceptual and organizational approaches will be foregrounded (acknowledged) in international standards? Will the different concepts and national practices combine to form new standards? This seems highly unlikely as concepts and practices differ, making them to a large extent incompatible. It can be assumed that the standards which are the most widely acknowledged and implemented on the international or national level will be used to develop the international standard in the domain in question. In the development process, competition may arise between influential national standards and standards backed by lobbies (such as ASIS in business continuity activity, for instance). For this reason, following the example of technical standards, corporate standards, called “management standards” are used by consultants in the domain and by companies aiming to apply for certification of their activities.

Lastly, the intertextuality and performative dimension of standards, which play a crucial role in their application, remain largely unexplored.

The formatting constraints imposed by standards create the conditions of a «performative» [2] that is to say a form of action is present, both intrinsically and extrinsically, in the writing process of these formats.

Hence the importance of:

- monitoring standards during the writing process;
- perusing and comparing standards closely related with the domain under the process of standardisation within different technical committees;
- confronting and attempting to ensure agreement on the definition of terms and concepts.

These three issues are generic and trans-sectorial. In the NOTSEG project, we have chosen BCA (Business Continuity Activity) as the field of experimentation for our methodology but the tools designed are expected to be applicable to other fields.

3. HYPOTHESIS AND METHODOLOGICAL APPROACH

Theoretically, research in progress deals with the relation between media flow of knowledge and natural language processing, in particular text mining tools.

The qualitative analysis of textual data enables data to be processed in such a way as to reveal the heterogeneity of large text corpora. Information traceability places communities (in terms of professional cultures, lobbies, etc.) and memory at the heart of knowledge flow, thus highlighting the interest of a socio-cognitive and political approach to these issues. Grasping and elucidating the diversity of points of view (political, economical, institutional, national, etc.) of stakeholders, who express their opinions in various spatio-temporal and cultural registers, is a tremendous task.

Our proposal is to combine natural language processing and linguistics with the sociology of science, in order to produce a fruitful analysis of corpora [3], since the heterogeneity of the texts imposes complex computing and epistemological processing.

Using concepts from text linguistics [4] is mandatory if one is to identify enunciative polyphony and to build classifications or cartographies. When combined with insights from the sociology of science [5] concerning stakeholders, their institutions and industrial or scientific policies, these representations could provide relevant categorizations of knowledge flows or controversies².

In the context of the NOTSEG project, we focus on the role of text mining tools which can provide a new way to apprehend the complexity involved in comparing large numbers of texts. This kind of approach combines linguistic engineering, knowledge engineering and knowledge communication [6].

In the first stage, we aim to check a methodological hypothesis concerning the comprehensive analysis of a corpus of standards by using qualitative text mining tools. More precisely, we attempt to identify the diversity of stakeholders and their institutions by analysing the tracks they have left in the text.

Characterizing stakeholders profiles.

Presently, we have identified several profiles of stakeholders who are likely to provide, prescribe or use standards. A diverse range of practitioners is involved:

- editors of standards: their role is to edit standards by comparing them with other existing standards on closely related topics produced by other NGOs or national, European, or international bodies.
- prescribers and their representatives (companies): these are officials (employees or consultants) in charge of making use of a standard in organizations (public or private) and monitoring the impact of these standards on legislation and regulations.
- stakeholders involved in making use of these documents and in the development of the corresponding operations.
- end users and their representatives.

² Massachusetts Institute of Technology, « Mapping Controversies », [www.demoscience.org/resources/index.php]

Socio-organizational modelling of actors

This modelling step aims at making the connection between standards and their socio-organizational context based on criteria related to geographical and institutional origins and on working group participants. This categorization has been implemented thanks to the work of AFNOR which carried out a cartography of the people involved.

Institutions include international standardization bodies, European or national standardization bodies, representatives of States (Ministries), NGOs, private actors (from industry and French or European think tanks). Participants in working groups - the editor and expert-members - are also taken into account, as are editors and expert-members belonging to companies. Figure 1 shows the different elements of this modelling in the form of a UML diagram of classes.

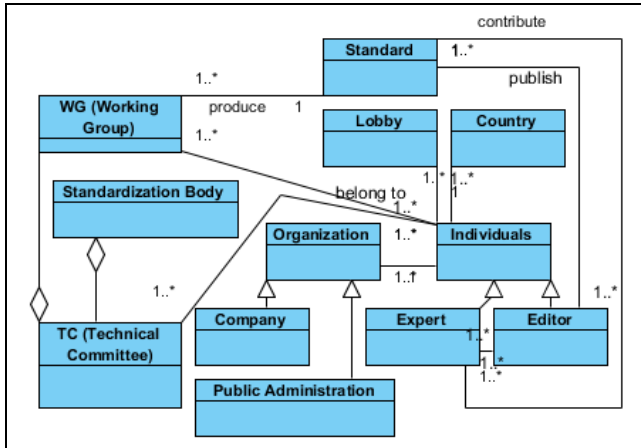


Figure 1: Draft representation of the textual, conceptual and semantic context (UML formalism)

Construction of lexical maps : contrastive glossary of lexical or semantic variations

Undertaking semantic and pragmatic analysis should make it possible to build a dynamic glossary (in English and French) in the field of security – limited first to a sub-field of security which accounts for semantic variations of terms and variations in their uses. This glossary is composed of a list of terms systematically defined in standards in the section “terms and definitions”.

This work does not aim at imposing a particular point of view or a unified vision of the field investigated. Conducted according to a comparative approach, this semantic and pragmatic analysis of key-terms in the studied standards aims at revealing zones of convergence, divergence, and cross-checking, including controversies, based on criteria which have to be identified. This glossary should account for semantic variations in terms and variations in their contexts of use. To achieve this aim, several possible methods of work are being investigated:

- the hypothesis of a semantic corpus search with the help of text analysis software (for instance, the list of terms defined in the section “terms and definitions” of the standard).
- expert evaluation concerning the identification of key-terms (corresponding to an interpretative choice of terms),
- a combination of the two approaches.

Concerning the semantic search hypothesis, the techniques and practices related to “clusterisation” (linked to the aggregates

method, cf. figure 3) and to graphic – particularly cartographic – representations which derive from it, could enable the construction of networks of categories and concepts, or networks of co-citations to identify actor networks.

Centered on the comprehensive analysis of textual data, computerized qualitative analysis tools can account for the studied corpus in a detailed way and yield a representation in the form of a graph (through tools such as Alceste, Lexico3, or WordMapper, for instance). The construction of networks of associated words (cf. figure 2) is among the techniques implemented by these tools.

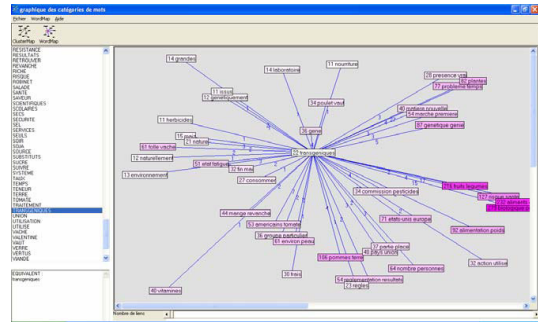


Figure 2: Display of the context of use for a word in WordMapper

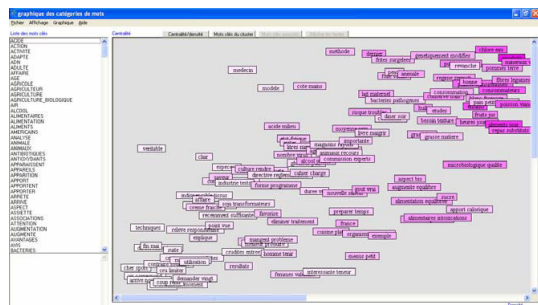


Figure 3: Diagram of clusters in WordMapper

In national, European and international standards, the principle is to analyze documents by comparing language expressions at the level of words, nominal groups, phrases, and paragraphs. The objective is to understand the meaning of identical language expressions in the international standard and in one or several national or European standards, or, on the contrary, their absence or their relative scarcity.

This work is an attempt to go beyond a debate of opinions and to identify a method which provides metrics for comparing texts, i.e. which makes use of statistical results calculated on the basis of the presence of linguistic markers. In other words, it is possible to conduct an analysis in terms of frequency of occurrence or to display the terms which frequently co-occur in a given standard, for instance, thereby displaying the context of use of a key-word and describing its semantic environment. Another issue is to calculate, in a standard, clusters which are built according to indexes of centrality and density. We hypothesize that this cartography would reveal, for a given standard, the semantic environment of terms considered as key-terms (with the option of returning to the text from the diagram clusters in order to carry out a more in-depth analysis).

4. EXPERIMENTATION

Construction of local ontologies in the form of graphic and interactive local ontologies

First, we propose to put at the user's disposal tools which offer a synthesis of the standard's contents in a cartographic or textual form. These new representations are designed to help the appropriation of concepts, ways of thinking, ways of organizing or defining modes of operation, these elements being connected to their socio-organizational context.

This work will lead to the construction of a plurality of ontologies – each of them representing a point of view – which will take into account the main “local ontologies” of national, European and international standards.

The originality of this approach is the construction of local ontologies which represent points of view corresponding to semantic worlds of standards studied individually. While an ontology is often used to produce a unified representation of a field, here it is used to express the point of view of a category of stakeholders and a sector of activity.

The methodology consists in constructing a conceptual and lexical representation for each standard, described for example in terms of concepts, organization, practices and economic competition. This work implies the design of a socio-technical device (section 5) with the following characteristics:

- This device, which makes it possible to group or contrast concepts, takes the form of an ontology.
- It is a formal language for the description of concepts and their inter-relations. This ontology is built according to a pragmatic approach, both linguistic and non linguistic, introducing parameters such as actors, editors, committees, nationalities or contexts (industrial or cultural).
- This technical device uses the semantic variations present in terms associated to texts.

Although an ontology is always a more or less objective representation, the construction of an ontology is an essential step in building a conceptual representation. An ontology describes a domain with concepts which are theoretical schemata which favour the intelligibility of phenomena.

Comparative analysis of local ontologies

Then, a comparative analysis of the network of concepts built will be conducted in order to build bridges between them. This analysis will rely on a device which allows navigation between different linguistic expressions of the same concept in all the texts. Some visual support will be provided, with a set of colours for instance, to assist the user.

The goal is to launch the construction of interactive cartographies which facilitate cross-comprehension and provide comparative information about texts of standards and their associated context.

These cartographies will cast light on overlapping, similar, different or convergent text areas between national or international standards.

Another goal is to construct a general ontology from these local ontologies. The ability to identify an abstract concept related to local concepts and their linguistic descriptions in texts of standards would be very useful for monitoring standards.

In order to assess our methodology, we have undertaken experimentation in the field of BCA. This experimentation comprises the following steps :

1. Corpus selection: selection of standards in a more restricted field than that of information security. We have selected BCA;
2. socio-organizational inventory of the universe of standardization, in connection with national bodies (AFNOR);
3. Building a glossary: Comparative analysis of texts of standards;
4. Building the conceptual map and local ontologies; setting up a referential of terms with the help of experts; identifying words which are important for the description of professional knowledge or practices;
5. Specification of the platform.

In this paper, only steps 1 and 5 are described.

Corpus selection

Within the large corpus of standards on global security management, we have chosen the topic of BCA and risk management because we are involved in Working Group 4 of the ISO Technical Committee TC 223 on societal security. Among the various standards that have been published in different countries (USA, UK, Australia, France), the following have been selected for the present study:

Standard reference	Title
BS 25999-1:2006	Code of practice for business continuity management
BS 25999-2:2007	Specification for business continuity management
ASIS SPC.1-2009	Organizational Resilience : Security, Preparedness, and Continuity Management Systems – Requirements with Guidance For Use
NFPA 1600. 2010	Standard on Disaster/Emergency Management and Programs.
BCI Good Practice Guidelines 2010	A Management Guide to Implementing Global Good Practice in Business Continuity Management
AS/NZS 5050:2010	Business continuity – Managing disruption-related risks, Standards Australia
ISO FCD 22301:2010	Societal security
ISO WD 22399:2010	Guideline for incident preparedness and operational continuity management,

ISO/CEI FDIS 27031	Information technology - - Security techniques -- Specification for ICT Readiness for Business Continuity (FDIS) SC27
ISO/CEI 24762:2008	Technical securities
NF ISO 31000 (ISO/IEC Guide 73)	Risk management - Vocabulary. TMB
IEC/ISO 31010	Risk management – Risk assessment techniques
Draft ISO/IEC	Guide 81 -- Guidelines for the inclusion of security aspects in standards.
NF ISO/IEC IS 27005	Information technology - - Security techniques -- Information security risk management. SC27

5. PLATFORM SPECIFICATION

One of goals of the NOTSEG project is to specify a platform dedicated to the management of a large corpus of standards, between 15 and 100 texts, applied to an engineering field, such as for instance BCA. This platform, designed to provide a common environment, will be used by two kinds of end-users: first, by consultants and editors of standards to help them during the writing process of a new standard (cf. section 3.1); second, by corporate departments in charge of tackling the implementation of standards in order to comply with national or international regulations (cf. 3.2). Two main ideas underlie our approach: one is to provide cartographic representations of standards, the other to assist the navigation of an end-user through the corpus of standards.

Considering that there is no universal representation independent of the goals and the organizational context, we aim to provide both graphic and textual representations, and several tools enabling comparison between several standards. It must be emphasized that all the representations are interconnected and that the platform provides specific interfaces allowing the end-user to navigate between them. Furthermore, this navigation will be assisted by applying specific knowledge based on the NaviText model [7].

Textual and Graphic Representation

The glossary of the standardized domain is the main textual tool. For each term, semantic and usage variations in the selected corpus are provided and enriched links (see section 5.2 below) can be followed; at any moment, the textual contexts of the same term in two standards can be compared by accessing them in one or several standards. This very simple tool is extremely useful to preserve conceptual coherence during the writing process of a new standard by using the same word to refer to an identical concept or on the contrary by choosing a new word to highlight the creation of a new concept.

Graphic representations complete the glossary. As explained in section 3.5, we consider that conceptual maps (or local ontologies) provide a useful level of abstraction, while at the

same time keeping and foregrounding the relations between concepts and qualifying their semantics.

For example, relations could be linked to the different phases of the PDCA cycle which governs all the standards.

Assisted Navigation

One of drawbacks of printed standards is that few tools exist for navigating through them, such as indexes and tables of contents. Thanks to digitization, a wide range of possibilities are presently available. The main point is the granularity of the objects accessed. In the NOTSEG platform, the basic granularity is the word. From a structural point of view, words are included in one or more sentences, which are included in a paragraph, which are included in a section and so on. From a text point of view, the frequency and context of a word, and co-occurrence networks, can be computed and linked. From a semantic point of view, the definition of a term in the standard, and the semantics of verbs which co-occur with specific terms, can be automatically annotated with metadata. Finally, from a pragmatic point of view, organizational metadata can be manually added in order to highlight the influence of certain lobbies on the definition of concepts.

All these data and metadata will be used by navigational knowledge [8] to afford different means of circulating through the standards. This kind of tool should enhance the intelligibility and comprehension of standards for engineers responsible for implementing them in companies.

Technical issues

The platform must be interoperable and in line with international standards of knowledge representation (RDF, OWL), and offer API in order to cooperate with other software. Two paths are checked. First, Protegé designed by Stanford University (<http://protege.stanford.edu/doc/>) is used worldwide and provides a sturdy background to implement the ontology. Furthermore, an endpoint SPARQL is included to store data and metadata.

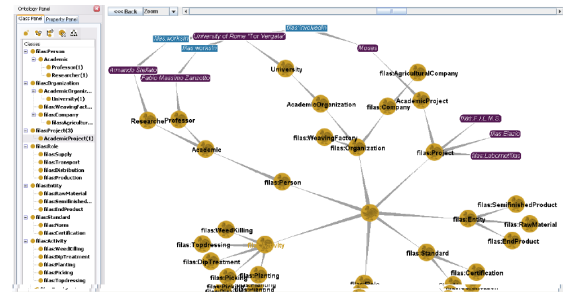


Figure 4: A screenshot from Semantic Turkey [9]

A second option is Semantic Turkey (ST) designed by Tore Vergata University [9], which is a “a Firefox based Knowledge Management and Acquisition Platform for the Semantic Web”. The main asset of ST is its ability to combine texts and several ontologies and to provide tools to keep track of concepts and their usages in different texts (see figure 4). Furthermore, an experimentation in the legal domain is in progress at the Artificial Intelligence Department of Roma 2.

6. CONCLUSION

We have described a methodology based on hypotheses of how to combine natural language processing tools with a sociology of science approach.

We have presented the specifications of a platform which is dedicated to editors, prescribers or end-users in the field of standardization and work in progress in the domain of BCA.

This specific study on the texts of BCA standards will be used for a more generic task, as the empirical core of a wider-ranging, prospectively designed inquiry.

The first step is to identify the main concepts and recommendations (operational, behavioral, technical, etc.) in which these formats are embodied. In the second step, this work on texts, related to the socio-organizational context of their production and communication, should provide information to identify or critically assess the new technopolitics already developed or under development in the field of security and crisis management.

Finally, this work will provide the opportunity to begin exploring, in the industrial domain, ways of thinking about the culture of security and risks which are presently the domain of engineering.

7. REFERENCES

- [1] L. Thévenot., "Un gouvernement par les normes. Pratiques et politiques des formations d'information", in Conein B., Thévenot L. (eds), "Cognition et information en société", **Raisons Pratiques**, No 8, 1997.
- [2] J. L Austin., **Quand dire, c'est faire** (original title: **How to Do Things with Words**), 1962, trad. fr. 1970, Paris, Seuil, coll. Points essais, 1991
- [3] D. Demazière., C. Brossard, P. Trabal, K. van Meter (eds), **Analyses textuelles en sociologie, logiciels, méthodes, usages**, PUR, 2006.
- [4] J.-M Adam., **Les textes, types et prototypes : récit, description, argumentation, explication et dialogue**, Paris, Nathan, 1992.
- [5] B. Latour **Changer de société. Refaire de la sociologie** (Titre original : Re-assembling The Social. An Introduction To Actor-Network Theory, 2005), Paris, Ed. La Découverte, 2006.
- [6] B. Juanals B., **La circulation médiatique des savoirs dans les sociétés contemporaines**, Université Paris Diderot-Paris 7, 2008.
- [7] J. Couto J., J.-L Minel., "NaviTexte, a Text Navigation Tool", **Lecture Notes in Artificial Intelligence 4733**, Springer-Verlag, p. 251-259, 2007.
- [8] J. Couto J., J.-L Minel., "A Linguistic and Navigational Knowledge Approach to Text navigation", **IJCNLP**, India, 2008.
- [9] M.T. Paziienza, S. Sguera, A. Stellato; "Let's talk about our "being": A linguistic-based ontology framework for coordinating agents". (R. Ferrario, & L. Prévot, Eds.) **Applied Ontology, special issue on Formal Ontologies for Communicating Agents** , 2 (3-4), 305-332, 2007.