

A conceptual ontology in the water domain of knowledge to bridge the lexical semantics of stratified discursive strata

Jean-Louis Janin, PhD student, U. Bordeaux Montaigne (tureennejlj@orange.fr)
Henri Portine, Professor emeritus (henri.portine@u-bordeaux-montaigne.fr)

September 2017

Abstract

This paper illustrates the semantic articulation between pivot and satellite lexical units of different discursive strata in the water lexicon model supported by the French Water Academy (*Lexeau* project of a *Bilingual lexicon of water related texts and data*, to improve public awareness of water related issues and ensure a better inter-comprehension among stakeholders. The lexical treatment of the discursive unit "water withdrawal" into stratified entries shows the capacity of a domain ontology to set a bridge between discursive strata in different languages, making easier internal and external translations between stakeholders. With more than a hundred lexical entries tested, there is an opportunity for a consortium of realization and a project of a pilot internet application.

1 Introduction

This paper presents one aspect of the lexicon project launched by the French Water Academy to improve public awareness of water related issues and ensure a better inter-comprehension among professional and non professional stakeholders of the domain. The focus is put on the role of the ontology in articulating pivot and satellite stratified lexicon units entered in different discursive strata.

It has been recognized that the meaning of a discursive unit may vary in the domain of water discourse, depending on the audience (Payen (2013)). It may lead to ambiguities and misunderstanding if some words and expressions are received by the audience in a sense which was not foreseen by the speaker or the author, *i.e.* out of the acception he is used to employ and trying to convey in his own discursive stratum. According to the large part of legal and administrative texts in the water domain and the large number of scientists and engineers involved in research and development activities, including humanities, we have introduced two different *technico-scientific (TS)* and *technico-administrative (TA)* discursive strata, together with the *current (C)* discursive stratum, for daily exchanges and media production on water actuality. In everyday life, words are linked to casual and colloquial notions or concepts. We have also introduced a *decisional/incitative (DI)* discursive stratum to take into account other textual productions of the domain. One example is the UN General Assembly Resolution of 2010/07/28 on "The human right to water and sanitation", still to be enforced by normative laws in each member state. The two main features of the stratified lexicon are the one-to-one relationship between conceptual units and pivot lexical units and the ordering of the discursive strata to limit the articulation of pivot units with satellite units of a higher range. The strata are listed with their prefixed label and corpus summary:

- 1-TS Technico-scientific (scientific papers, articles and books);
- 2-DI Decisional incitative (political and legal incitative proposals and decisions);
- 3-TA Technico-administrative (laws, directives, regulation, judicial activity);
- 4-C Current (current exchanges and daily media : press, radio and television.)

This order corresponds to the intuitive idea that concepts of scientific origin are likely to be transferred in the lexicon as pivot lexical units, defined through object and data properties of the ontology and articulated with satellite units in discursive strata of a higher range. An articulation with discursive strata of a lower range would invalidate our assumption, but we could not find any example of it.

After presenting a partition of the water related domain of knowledge, we will present the *model of water movements and anthropic water flows* introduced in the ontology, and one example of a pivot unit articulated with three satellite units. The following discussion focuses on the central place of the ontology in the model device including the end-user. The question of proper names is discussed, together with the relations of our research work in linguistics to lexical semantics and terminology.

2 A partition of the water related domain of knowledge

The water related domain of knowledge has been divided into three parts. The first part, called ONTOLOGY, is an ontology of the concepts, mainly scientific and administrative, structuring human knowledge and activities in water use and aquatic environment conservation and restoration. The output of this part of knowledge is composed of conceptual units, documents and typed individuals (fig. 1).

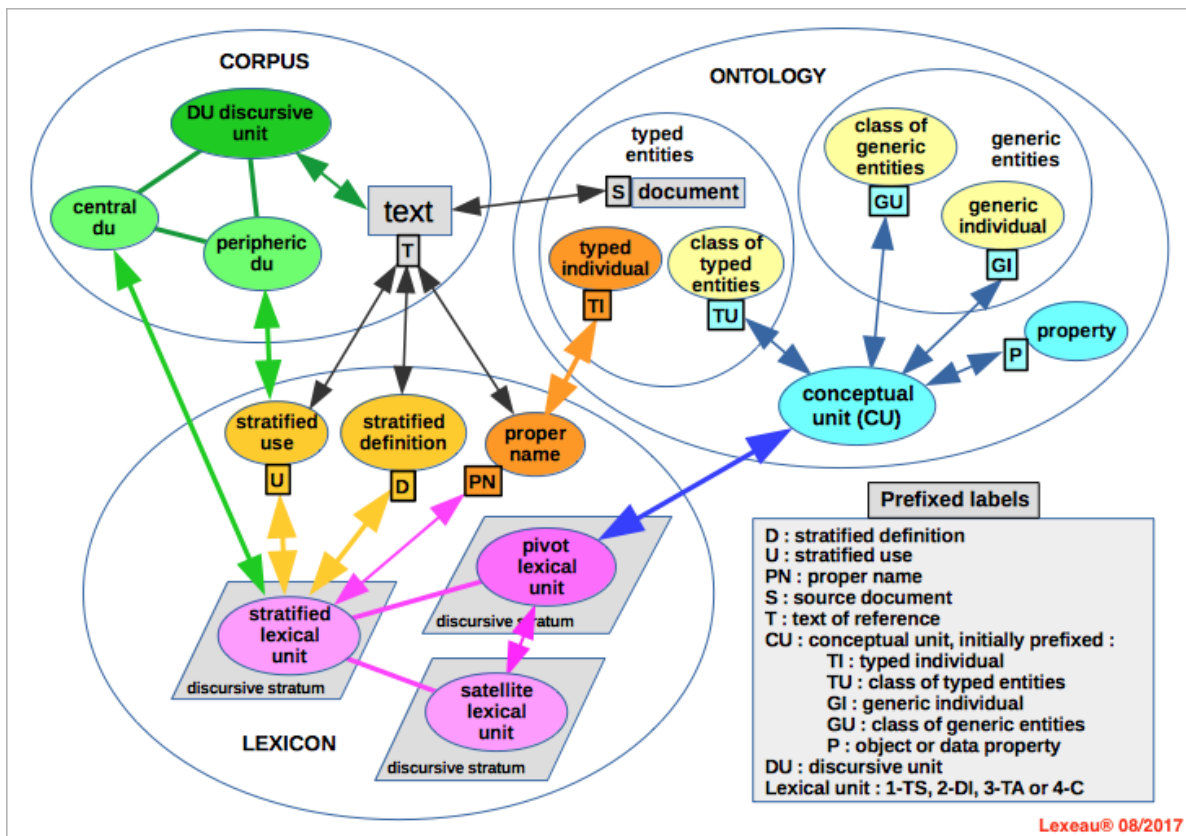


Figure 1: A partition of the water related domain of knowledge

The second part, called *LEXICON*, is composed of proper names and stratified lexical units. Pivot units have a *one-to-one* relationship with conceptual units and a *one-to-n* relationship with satellite units. The third part of the domain, called *CORPUS*, is composed of recurrent discursive units found in texts referenced in proper names implementation, lexical definitions and stratified examples of use of some lexicon entries. The relevance of the use of concepts — including scientific concepts, except the concepts of mathematics, physics and chemistry — is textual: hydrology, geology and so on are stated in texts. That is why corpus linguistics is of overall importance for ontologies. The central discursive units are linked with the lexical units and the peripheric ones are linked with the stratified examples of use.

3 A conceptual model of water movements and anthropic water flows

The water withdrawals are modeled as upstream water movements generating anthropic flows recovered by downstream water movements. The graph of the model is presented fig. 2. We have added, on a yellow background, the labels of the class relations (*object relations*) and, on an orange one, the labels of a set of related individuals : The *Water withdrawal1* generates the *Water flow (anthropic)1* which is recovered by the *Water restitution1*. The figure 3 presents the creation of the three conceptual units,

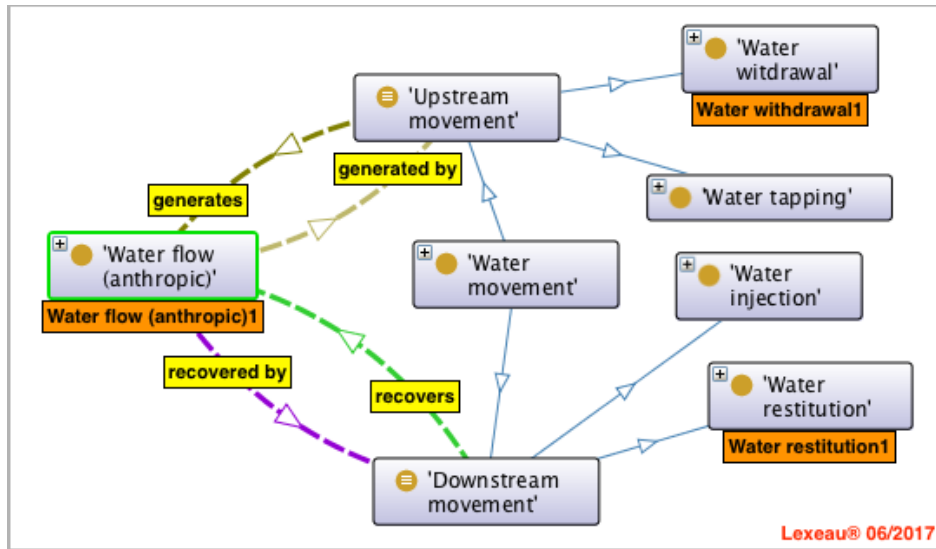


Figure 2: The model of water movements and anthropic flows with a set of related individuals

prefixed CU, after creating the three concepts, prefixed TU, as individuals in their original labelled class, further added to the class *Conceptual unit* as individuals of the hyper-class *Typed entity*. To present the

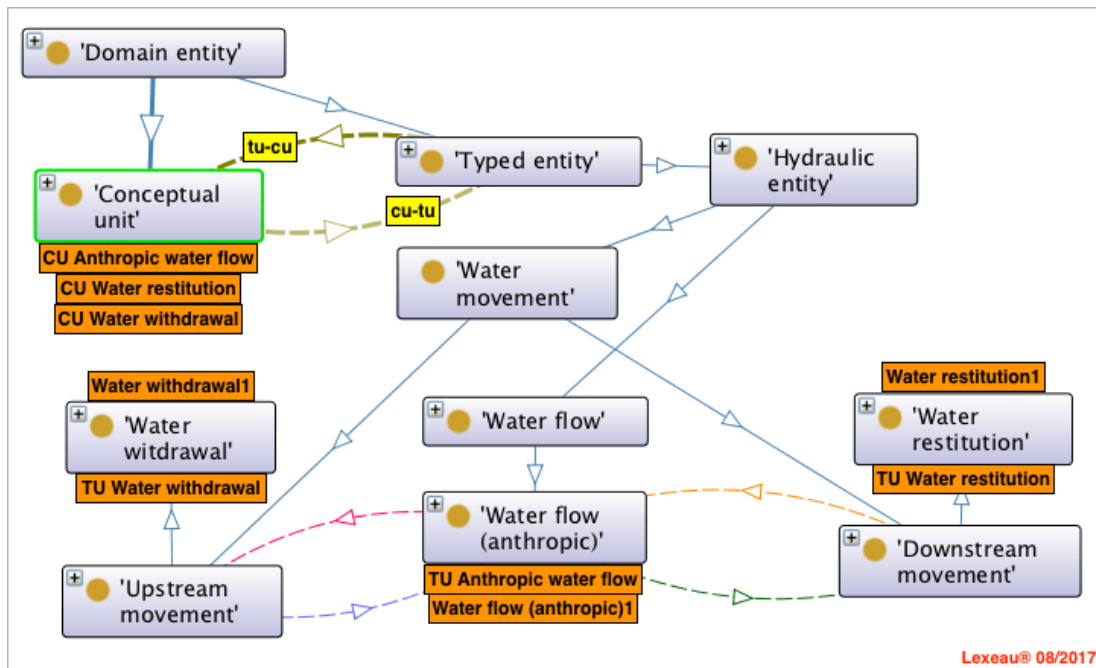


Figure 3: Generating three conceptual units through individuals created in labelled classes

lexicalisation process and text-based processes within and between the three parts of figure 1, we have used the same software (Protege 5.2) to edit a *project ontology* of all the entities of the partition.

4 Articulating pivot and satellite lexical units

A graph of the *Conceptual unit* and *Stratified lexical unit* related classes and of the four sub-classes of lexical units is presented figure 4, with the pivot-satellite articulations between these units. The four related labelled lexicon units and their labelled relations have been added on orange and rosy backgrounds.

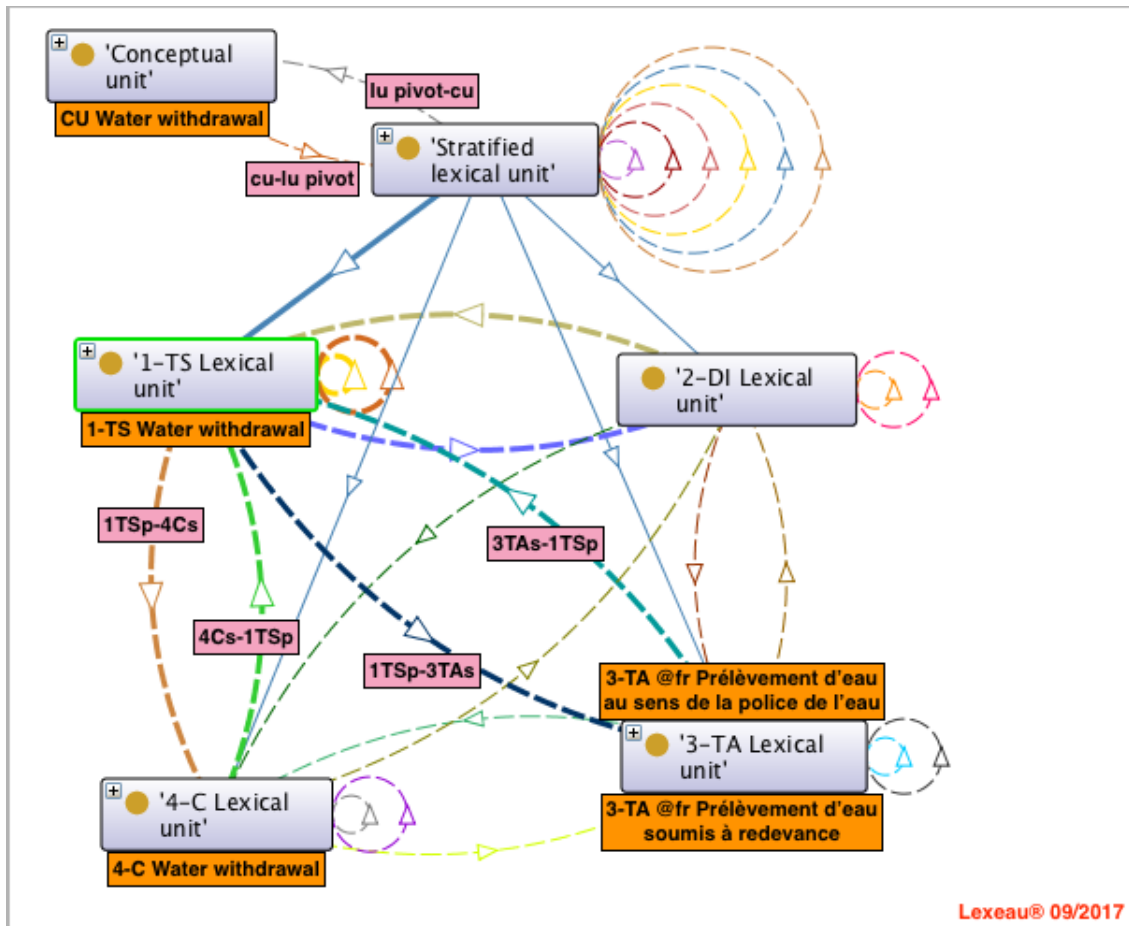


Figure 4: An example of one pivot lexical unit articulated with three satellite units

The definitions of the pivot units are derived from the object and data properties of the ontology entities whose labels are saved as concepts and further lexicalised (cf. fig. 2 and 3), leading to :

- D 1-TS Water withdrawal : *A water extraction carried out by an operator in a water body which generates an intermittent or permanent water flow. The operation can be associated with an extraction water work equipped with a volumetric meter*
- D 1-TS Anthropic water flow : *Water flow permanent or intermittent over a given period, in a water use concern with a physical or legal person. With a total volume and a maximum flow rate measured or estimated, it is associated with an upstream water movement (water withdrawal in a water body or water tapping in a water network) and a downstream water movement (water restitution in a water body or water injection in a water network)*

The construction of the lexicon is based on the articulations of pivot lexical units with satellite units, looking for other acceptations of the same word or a closely related form in a different discursive stratum. An example is presented figure 4 with the pivot unit *1-TS Water withdrawal* and the satellite units *4-C Water withdrawal*. Two entries of *water withdrawal* are found in the French administrative discourse, with two different meanings referring to the "police de l'eau" and "redevances de l'agence de l'eau"

national practices with no exact equivalent in Great Britain, at least to our knowledge. The entries are defined in French and just labelled with the “@fr” prefix in the *Anglo-british* version. (fig. 5 and 6).

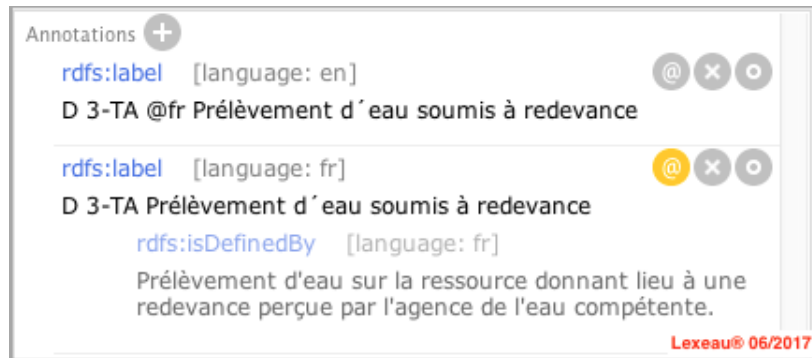


Figure 5: "3-TA Water withdrawal" first acceptance in France



Figure 6: "3-TA Water withdrawal" second acceptance in France

To finalize the English version, possible administrative acceptations of *Water withdrawal* should be checked in the United States, Canada, Australia, New Zealand and other English speaking countries.

Looking for a satellite lexical unit in a current bilingual discourse, we have found it in a plural form, on the OECD Data site on water withdrawals, with a bilingual definition (fig. 7).



Figure 7: A bilingual OECD definition of "4-C Water withdrawal"

The lexical unit *4-C Water withdrawal* is defined as a volume of water withdrawn yearly for human

use, except for hydroelectricity production. It is expressed in millions of cubic meters per country on the tables and maps of the OECD site. This meaning is easier to understand than considering the *hydraulic operation* modeled in the ontology, with its related entities, or the *administrative operation* authorized by the French *police de l'eau* or taxed by the French *Agence de l'eau* under specific circumstances.

With two different acceptations, the French administration as a whole is not prepared to match individual data collected by the "police de l'eau" and the "Agence de l'eau" (operator, volume of water withdrawn, final user, etc.) as the practice on both sides do not state how these data on water withdrawals should match in the *real world*. One way to overcome these difficulties would be to get more information on both sides, typically about the downstream operation on the flow of the water withdrawn, separating the notions of water movement and water flow, as proposed in the definitions of *1-TS Water withdrawal* and *1-TS Anthropic water flow*, in a scientific approach of the hydraulic concepts involved.

5 Discussion

5.1 A conceptual ontology at the heart of the project

Bringing a conceptual ontology within the construction process of the water lexicon raises the question of its feasibility and its maintenance among a community of contributors, mainly scientists, engineers and jurists, in such a vast domain of knowledge. The main issue of the project is to find out, in public and private enterprises, trained contributors to update the ontology on the long term in agreement with their employer. Existing ontologies or relational data bases have already been developed successfully for specific uses in the industry of water services or scientific laboratories, with specific concepts and terms involved in their internal use. The aim of the project is by no mean to plunder existing software or replace it by a new one. In the modeling of water movements, the objective was to organize existing and new concepts and label them on this particular theme so they could be shared and traced in published texts, in English and/or in French, and further introduced as lexicon entries. With no direct users, the ontology may contribute to share the labels of scientific entities and facilitate inter-comprehension between scientists and engineers, and further institutional data exchanges. The place of the ontology in the device model is presented in figure 8. Another issue is the sharing of concepts with energy and food

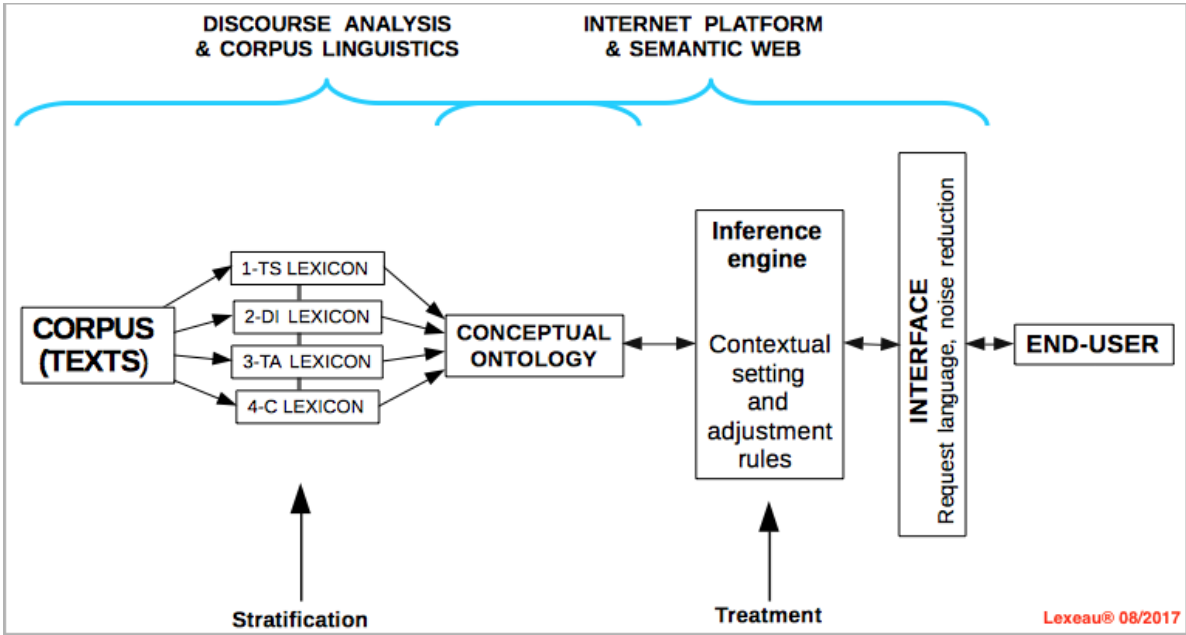


Figure 8: An overall model of the device

related knowledge domains. This issue is addressed with the notion of energy, water and food *security*

nexus pointed out by Bazilian et al. (2011) and updated by Howarth and Monasterolo (2016). It should take place through the study of an ontology alignment in the common domain of knowledge and activity.

5.2 The question of proper names

Proper names — as proper nouns, including acronyms, and nominal phrases — are of considerable importance in the domain of water. They are used to name places, persons, documents, events, etc. We follow the position of Saul Kripke on the nature of proper names as rigid designators, against the descriptivist theory of Russell and Frege, assuming their identity as a collection of finite descriptions (Kripke (1982), Kripke (1980)). In our project, nevertheless, proper names are issued from the individuals of the typed classes of the domain entities. The class object property and data property assertions of these individuals correspond, when they are phrased, to the finite descriptions of the descriptive theory. The proper names introduced in the lexicon are not stratified, with no use to define them as it would duplicate the properties of their homologue individual in the ontology.

5.3 The relation to lexical semantics

Our relation to lexical semantics in the project follows the historical evolution of the discipline. At the start, lexicology tried to assess the sense of words, considered as isolated entities, within a network of synonyms and antonyms (cf. Cadiot and Habert (1997), Panier and Rémi-Giraud (2003), Victorri and Fuchs (1996)). Further relations appeared later : hyperonymy *vs* hyponymy, then meronymy, and the question of holonymy. The research focus moved then on cooccurrences and collocations, leaving the domain of the auto centered lexicon for a corpus-based lexicon driven by discursive activities. Our research work takes into account the discourse and the social activity where it takes place (small talk, commercial, administrative, educational, scientific), which raises the question of our connection with sociolinguistics, when we infer a stratified lexicon, depending on the audience (cf. Labov (1972), on speech style).

In its bilingual aspect, our work is in debt with English textbooks (Murphy (2010), Cruse (1986)), including a grammar (Huddleston (1988)). Handbooks did help to get some insight on the processing of language by computers and knowledge-based systems (Mitkov (2003), Baader et al. (2007) and we are concerned with recent developments on *Big Data* (Bidoit and Doucet (2017)). Our work is confronted and in debt with classical linguistic results about conceptual analysis (Wierzbicka (1985)), sense-text linguistics (Mel'čuk and Polguère (2007)), the generative lexicon (Pustejovsky (1995)) and *inferential* lexical semantics (Kayser (1997a), Kayser (1997b)), as proposed in the abstract of the last reference ¹ :

Actually, what matters [in lexical semantics] are the inferences which are warranted by the use of a word in a context; the existence of a referent in some universe of discourse is merely a possible by-product of these inferences.

The discursive affiliation of a lexicon entry depends on the professional affiliation of the speaker and the inferences that he can make and share with his audience. This leads to different acceptions of the same word and implies different combinations with other words in the text. The journalist having in mind a *water withdrawal* as a volume of water withdrawn in one year will not make any distinction between this water withdrawal and the corresponding yearly *water flow*. He would not understand what *water restitution* stands for, without a clear view of the model presented here. Same thing for the administrative authority licensing a given operator for a given water withdrawal, with no clear distinction between the operator and the end user and no information on the stage of use (final *vs* not final) and the level of use (collective *vs* private) of the water withdrawn.

¹http://www.persee.fr/doc/lfr_0023-8368_1997_num_113_1_5372

The use of current expressions may result in considerable ambiguities and erroneous conclusions. As pointed out by G. Payen [Payen (2013) p.113], the current discourse does not always make the difference between a problem of a *water shortage* due to the lack of water resource and a situation of *water shortage* in an urban district out of reach of the municipal drinking water network of the town.

5.4 The relation to terminology

Terminology is above all technical and scientific. It has little interest in social relations. It is a normative activity, to facilitate the integration in a native language of words and expressions already used in a foreign second language, most of them in English, with a highly controlled definition in the source language. The terminologist tries to accommodate the original definition in his native language. To fulfill business and trade issues, the activity is *prescriptive*, looking for a one-to-one relation between the terms of each language to facilitate the translation of documents. Linguists dealing with terminology are well aware of the difficulties of their task (Depecker (2003), Condamines (2016)).

We are engaged in a *descriptive* corpus sourced activity, in French and in English, with no prescriptive pretension and the clear notion that conceptual units may not have their equivalent in two languages, which can be rendered in the bilingual ontology and the lexicon entries. The tools used to scrutinize the thematic corpora of the project for recurrent discursive units will be able to analyse the use of prescribed terms in a given context, therefore contributing to some terminological activity.

6 Conclusion

After presenting the ontology of the Lexeau project in previous papers (Janin and Portine (2016) and Janin (2016)), this paper has presented, through the example of a model of water movements and anthropic water flows, the role of a conceptual ontology in the water related domain of knowledge and the sense of its threefold partition. In our example, the class object and data properties of the ontology are used to create stable definitions, in the technico-scientific discursive stratum, of the pivot lexical units *1-TS Water withdrawal* and *1-TS Anthropic water flow*. It allows to create other lexicon entries as satellite units to the pivot units, such as *4-C Water withdrawal*, in the current discursive stratum. An overall model of the device with its different modules has been presented. The treatment of proper names has been discussed, together with the relations of our research work to lexical semantics and terminology.

Intellectual activities are not limited to research activities, describing scientific phenomena. They have their place in professional activities of the water domain, aiming at a sound technical, commercial, ecological and legal water management. They have also their place in daily activities, in current exchanges and in the media production on water use, water pollution and security matters (drought and floods). These activities share the same words in different meanings, defined in different discursive strata. Through the example of *water withdrawal*, we have shown that the lexical units in the administrative and current discourse are underspecified in terms of the *real world objects* and *persons*. Their articulation with a pivot lexical unit with complete specifications is crucial for a better understanding of the stakeholders of water uses and withdrawals. The solution of internal *translation* difficulties in one language applies also when dealing with two languages, with legal and political impacts in international affairs.

With more than a hundred discursive units tested on the model device, the construction of an ontology combined with textual investigations appears to be a sound way of bridging current and specialized discourses and a promising contribution to mutual inter-comprehension of stake holders in the water domain. There is now an opportunity to shape a consortium of realization of a pilot project (an internet application). The goal would be to develop and test the application on thematic priorities, with minimal development expenses and the help of professional contributors and non professional end users.

References

- Baader, F., D. Calvanese, D. L. McGuinness, D. Nardi, and P.-S. P. F. (2007). *The Description Logic Handbook*. Cambridge: Cambridge University Press.
- Bazilian, M., R. Holger, M. Howells, S. Hermann, D. Arent, D. Gielen, P. Steduto, A. Mueller, P. Komor, R. S. Tol, and K. K. Yumkella (2011). Considering the energy, water and food nexus: Towards an integrated modelling approach. *Energy Policy (Elsevier)* 39, 7896–7906.
- Bidoit, N. and A. Doucet (2017). *Les Big Data à découvert*, Chapter La diversité des données, pp. 24–25. Paris: CNRS éditions.
- Cadiot, P. and B. Habert (1997). *Aux sources de la polysémie nominale*. Larousse.
- Condamines, A. (2016). Vers la définition de nouvelles langues contrôlées pour limiter le "risque langagier". In LIMSI (Ed.), *Atelier risques et TAL (JEP-TALN- RECITAL 2016)*, Paris.
- Cruse, D. (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
- Depecker, L. (2003). *Entre signe et concept, éléments de terminologie générale*. Presse Sorbonne Nouvelle.
- Howarth, C. and I. Monasterolo (2016). Understanding barriers to decision making in the UK energy-food-water nexus: The added value of interdisciplinary approaches. *Environmental Science and Policy (Elsevier)* 61, 53–60.
- Huddleston, R. (1988). *English grammar : an outline*. Cambridge: Cambridge University Press.
- Janin, J.-L. (2016). Sélection et typage des acronymes de PREVIRISQ INONDATIONS 2016. In AFEPTB (Ed.), *Les Ateliers PREVIRISQ INONDATIONS - 17 octobre 2016 - Centre des Congrès de Lyon*.
- Janin, J.-L. and H. Portine (2016). L'ontologie du projet lexeau®. In Université de Bordeaux (Ed.), *Actes JFO 2016 (ISBN 978-2-9558601-0-6)*.
- Kayser, D. (1997a). *La représentation des connaissances*. Paris: Hermès.
- Kayser, D. (1997b). La sémantique lexicale est d'abord inférentielle. *Langue française Aux sources de la polysémie nominale*(113), 92–106.
- Kripke, S. (1980). *Naming and necessity*. Cambridge (Mass.): Harvard University Press.
- Kripke, S. (1982). *La logique des noms propres*. Les éditions de minuit (Naming and necessity).
- Labov, W. (1972). *Sociolinguistics*, Chapter The Study of Language in its Social Context, pp. 180–202. Penguin Books.
- Mel'čuk, I. and A. Polguère (2007). *Lexique actif du français*. Bruxelles: De Boeck.
- Mitkov, R. (2003). *The Oxford handbook of computational linguistics*. Oxford: Oxford University Press.
- Murphy, M. L. (2010). *Lexical meaning*. Cambridge: Cambridge University Press.
- Panier, L. and S. Rémi-Giraud (2003). *La polysémie ou l'empire des sens : lexicque, discours, représentations*. Lyon: Presses universitaires de Lyon.
- Payen, G. (2013). *De l'eau pour tous ! Abandonner les idées reçues, affronter les réalités*. Paris: Armand Colin.

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge (Mass.): MIT Press.

Victorri, B. and C. Fuchs (1996). *La polysémie : Construction dynamique du sens*. Paris: Hermès.

Wierzbicka, A. (1985). *Lexicography and conceptual analysis*. Ann Arbor: Karoma.