

Le dictionnaire de l'Académie de l'Eau

Un lexique stratifié des données textuelles et géophysiques sur l'eau*

Jean-Louis JANIN**
Henri PORTINE***

3 décembre 2019

*Version révisée de la note présentée à l'assemblée générale de l'Académie de l'Eau le 14/10/2019

**ICPEF honoraire, Docteur en linguistique (contact : jeanlouisjanin@gmail.com)

***Professeur émérite de linguistique - Université Bordeaux Montaigne

1 Introduction

Tous les domaines de savoir se heurtent à la question du langage : peut-on partager les mêmes concepts même lorsqu'ils correspondent à des dénominations différentes ? peut-on faire coïncider un concept quotidien issu de pratiques et un concept scientifique (au sein d'un réseau) ? Certains mots, apparemment partagés, peuvent ne pas renvoyer au même sens, peut-on alors opérer des ponts entre ces différents sens ou élaborer une « façon de se comprendre » ? Tous les acteurs de tous les domaines de connaissance le savent : pour bien agir ensemble, il faut se comprendre, et plus qu'à demi-mot. Établir les correspondances entre des lexiques professionnels (techniques, juridiques, etc.), un ou des lexiques scientifiques et des lexiques de la quotidienneté (agents de terrain, clients, journalistes et autres médiateurs) suppose ainsi des formes d'appariement et d'ajustement parfois complexes ; c'est pourquoi les lexiques bijectifs échouent. Or les populations veulent de plus en plus comprendre les motivations de l'action publique et les acteurs politiques, juridiques, médiateurs (dont les journalistes) souhaitent de plus en plus l'adhésion du grand public. C'est pourquoi il importe, aujourd'hui, de mettre en regard les différents usages effectifs. Pour ce faire, Jean-Louis Janin a défini la notion de strate discursive : une strate discursive caractérise un ensemble homogène d'acteurs, à la fois producteurs et auditoires de discours au sein d'un domaine déterminé (en l'occurrence, le domaine de l'eau et de sa gestion, question à la fois importante et sensible). À l'aide d'outils permettant une description et une évaluation fines, il a proposé une catégorisation de ces strates discursives qui permet de mettre en relations leurs lexiques et leurs spécificités afin de promouvoir l'intercompréhension entre ces différentes catégories d'acteurs. Mais pour assurer les interrelations entre ces strates discursives, il faut un pivot : ce sera une ontologie du domaine de connaissance de l'eau. L'objectif est d'élaborer ce système d'intercompréhension en français et entre le français et l'anglais afin de lui donner une dimension internationale.

Il ne s'agit donc pas d'un projet terminologique mais de faire « jouer » ensemble des lexiques ayant leurs spécificités, un(e) chercheur(e) ne s'exprimant pas comme un(e) ingénieur(e), un(e) ingénieur(e) ne s'exprimant pas comme les utilisateurs et les journalistes assurant leur rôle de médiateurs et tentant de faire s'entendre les différents acteurs. Les buts des uns et des autres ne sont pas identiques :

- les industriels visent l'optimisation de leurs activités (collecte et traitement des eaux usées, valorisation et recyclage, production d'eau potable, distribution) mais aussi la diffusion et la promotion de ces activités dans le champ francophone et à l'international ;
- les politiques doivent prendre des décisions et fournir des préconisations en connaissance de cause et donc dans une parfaite compréhension des enjeux et des intérêts des uns et des autres afin de mener une action équilibrée ;
- la justice nationale et internationale doit pouvoir rendre ses décisions en toute sérénité et donc en ayant clairement compris ce qui se jouait à travers les différents discours tenus de part et d'autre ;
- les journalistes doivent rendre les débats plus transparents en éclairant l'opinion et en faisant remonter les positions s'exprimant au sein de cette opinion ;
- le grand public doit bien comprendre les enjeux mais aussi ce qu'il paie en échange d'un service et l'évolution écologique sans céder aux émotions ou aux rejets rapides que peuvent provoquer les réseaux sociaux. Les défis sont donc nombreux. Le projet de Jean-Louis Janin est parrainé par l'Académie de l'Eau et devient donc l'un de ses projets sous le nom Lexeau®. Une thèse de doctorat a permis de détailler l'entreprise, d'en assurer

la faisabilité, de le modéliser sur un échantillon et doit maintenant permettre d'élaborer un prototype à réaliser en open source (licence GPL) pour tester collectivement une plateforme web de consultation du dictionnaire bilingue. On trouvera en annexe deux figures extraites de la thèse de Jean-Louis Janin : l'une présente le dispositif d'ensemble du dictionnaire ; l'autre donne la partition de la base de connaissance sous-jacente au dictionnaire.

2 Un lexique pour aider à résoudre les problèmes actuels

La question de l'eau se pose en 2019 dans une problématique qui a évolué depuis la fin du siècle dernier, avec comme nouveaux enjeux la répartition des usages de l'eau en période de pénurie (loi sur l'eau de 1992), une vision partagée de l'eau en tant que ressource et en tant que milieu, et l'objectif d'un « bon état » des masses d'eau (directive cadre sur l'eau de 2000 et loi sur l'eau de 2006). La question des inondations et des épisodes de canicule et de sécheresse se pose à nouveaux frais aux climatologues, en termes de probabilités, et aux décideurs, en matière de prévention et d'assurance. Les débats sur les économies d'eau, le droit à l'eau et à l'assainissement et les retenues d'eau sont relancés et donnent lieu à des affrontements. Ils doivent trouver des issues concrètes, à travers un compromis social et sociétal pour maîtriser les effets délétères du changement climatique d'origine anthropique, en plus des compromis à trouver en matière d'énergie et d'alimentation.

La mise au point de ces compromis passe par une compréhension partagée des systèmes techniques et par une entente entre les partenaires, où figure le grand public, qui est aussi le corps électoral. Elle repose sur une formulation technique des problèmes partagée par les scientifiques et les ingénieurs et sur une vision socio-politique des solutions qui débouche sur des décisions acceptables par les populations concernées. Sans cette double approche, il n'est pas possible de mener à bien des actions d'envergure au plan national et international. La question du langage est essentielle pour y parvenir, y compris dans le passage d'une langue à l'autre, notamment lorsqu'une des langues, ici l'anglais, joue le rôle de *lingua franca* dans une bonne partie du discours technico-scientifique sur l'eau à l'échelle mondiale et dans la préparation des directives européennes. Il est essentiel que la sémantique de ces discours en anglais soit analysée à l'aune de l'emploi de cette langue comme langue native, afin d'écartier le risque du « *globish* », pratiqué comme une langue seconde dégradée, à laquelle chacun confère la sémantique qui lui convient, ce qui multiplie les malentendus. Ce principe s'applique aussi à l'analyse des discours en français natif quel que soit le locuteur. Il peut conduire à des emplois différenciés d'une même langue dans différents pays, comme le Royaume Uni et les États Unis pour l'anglais ou la France et le Québec pour le français.

Prenant appui sur une thèse de linguistique soutenue à Bordeaux en début d'année, l'Académie de l'Eau est en mesure de proposer à ses partenaires la réalisation d'un prototype sur une plateforme internet pour relever le défi de l'inter-compréhension entre les acteurs du domaine de l'eau, avec trois strates discursives — technico-scientifique, technico-administrative et « courante » — pour distinguer entre les modes de compréhension d'un auditoire de professionnels ou des journalistes du quotidien et des réseaux sociaux. Il est fait appel à une ontologie du domaine (description des connaissances) dont les concepts, réputés stables, sont transposés dans des unités lexicales « pivots » qui peuvent être associées à des unités « satellites » dans

d'autres strates discursives. Par exemple, l'unité lexicale pivot « 1-TS Prélèvement d'eau » est rattaché à la strate discursive technico-scientifique. Elle est transposée depuis le nom d'une classe d'objets du monde dans l'ontologie du domaine, tout comme l'unité lexicale « 1-TS Flux d'eau anthropique ». La transposition dans le lexique des propriétés relationnelles de ces objets dans l'ontologie du domaine débouche sur une relation lexicale entre les deux unités : « 1-TS Prélèvement d'eau » <gènère> « 1-TS Flux d'eau anthropique ». L'unité lexicale pivot « 1-TS Prélèvement d'eau » a comme satellites l'unité lexicale « 2-TA Prélèvement d'eau », avec deux hyponymes (« 2-TA Prélèvement d'eau au sens de la police de l'eau » et « 2-TA Prélèvement d'eau soumis à redevance »), dans la strate discursive technico-administrative, et l'unité lexicale « 3-C Prélèvement d'eau », dans la strate discursive courante, où elle est entendue comme un volume d'eau. Les définitions et les exemples d'emploi du lexique sont tirés des textes référencés dans l'ontologie du domaine. La mise à jour du lexique s'effectue sur une base documentaire, par une analyse textométrique des documents référencés par thème et éventuellement par discipline qui fait émerger les unités linguistiques récurrentes. Ces unités sont confrontées au lexique existant, avec de nouveaux concepts transposés si nécessaire dans de nouvelles unités lexicales.

Les tests de la maquette du projet effectués au cours de la rédaction de la thèse sur plusieurs centaines d'entrées du lexique permettent de passer à la réalisation d'un prototype en faisant appel à un prestataire rémunéré (informaticiens), avec la constitution d'un groupe de projet et d'une équipe chargée de finaliser le cahier des charges et de tester le prototype.

3 Résumé de la thèse

L'objectif de la thèse était de modéliser un lexique « spécialisé » dans le domaine de l'eau, utilisable par les spécialistes du domaine, par le grand public et par les journalistes (en tant que médiateurs entre ces deux groupes), dans une problématique d'inter-compréhension entre les différents auditoires. Trois strates discursives (auditoires-types) — technico-scientifique, technico-administrative (gestion et aspects juridiques) et « courante » — ont été introduites en entrée du lexique pour distinguer entre trois modes de compréhension, chaque mode résultant du choix des mots et de leur agencement en discours : les strates discursives technico-scientifique et technico-administrative visent un auditoire de professionnels du domaine ; la strate discursive courante s'impose aux journalistes du quotidien et sur les réseaux sociaux. Le nombre de strates a varié au cours de la recherche, dans une problématique de divisibilité d'un ensemble d'éléments. Plus un ensemble est divisé, plus la recherche est fine, avec un algorithme de plus en plus complexe pour différencier les éléments. Moins on le divise, plus on facilite le traitement, avec le risque de ne plus traiter le vrai problème. Nous avons cherché à rendre opératoire cette sous-catégorisation lexicale sans la rendre stérile, pour donner un sens aux unités discursives dans une ou plusieurs strates, en tant qu'unités lexicales stratifiées.

Pour rendre compte des spécificités des discours scientifiques, juridiques ou administratifs sur l'eau, le lexique a été adossé à une ontologie des objets de connaissance du domaine, ontologie dont il procède pour une partie de ses entrées avec les unités lexicales « pivot ». Ces unités sont issues des concepts édités dans l'ontologie et transposés dans une strate discursive qui reflète le contexte d'élaboration du concept. Elles peuvent être associées à des unités lexicales

« satellites » dans d'autres strates discursives, avec des lemmes¹ identiques ou proches. Les propriétés des classes de l'ontologie génèrent dans le lexique les relations lexicales classiques du type « Aléa » <a comme hyponyme> « Crue » et « Crue » <est un hyponyme de> « Aléa », mais aussi d'un nouveau type, par exemple « Crue » <déclenche> « Inondation de rivière ». Ceci fait du réseau des entrées du lexique stratifié un réseau « cognitif » avec son métalangage et ses inférences potentielles. Les classes d'entités du monde de l'ontologie du domaine permettent d'associer un dictionnaire des noms propres aux noms des classes d'objets du monde transposés dans le lexique. Par exemple, le nom propre « Garonne » est associé à l'unité lexicale « 1-TS Cours d'eau » transposée dans la strate discursive technico-scientifique depuis le nom de la classe « Cours d'eau » de l'ontologie du domaine.

La raison d'être d'un lexique opérationnel de domaine est d'y retrouver des unités linguistiques récurrentes dans les textes du domaine. L'émergence de ces unités est obtenue par une analyse textométrique des unités récurrentes et des cooccurrences, effectuée sur les textes regroupés (le corpus). Elle débouche sur la consolidation du lexique existant et sur de nouvelles unités lexicales « pivot » issues de la mise à jour de l'ontologie du domaine.

Le double processus de construction du lexique a été testé sur le thème des prélèvements d'eau et sur celui des inondations, après modélisation des sous-domaines correspondants dans l'ontologie. La discussion d'une quatrième strate discursive pour rendre compte des discours artistiques, religieux et idéologiques sur l'eau a conduit à son abandon, au profit de l'introduction des thèmes correspondants dans l'ontologie du discours. Ceci ouvre la voie à une recherche en sociolinguistique sur l'apport d'un lexique stratifié aux recherches dans le domaine des STS (Sciences, technologies et société), domaine qui recoupe celui des *Science and technology studies* du monde anglo-saxon.

Les figures placées en annexe illustrent le processus de consultation et de mise à jour du lexique (1) et la partition de la base de connaissance à l'origine des trois ontologies du dispositif. Cette partition permet de distinguer entre les entités qui relèvent spécifiquement du domaine de connaissance de l'eau, des discours sur l'eau et du lexique de l'eau et de faire apparaître les relations entre les sous-bases.

1. Forme canonique d'une entrée dans un dictionnaire de langue, soit, en français, l'infinitif pour un verbe et le masculin singulier pour un adjectif ou un nom.

A Figures

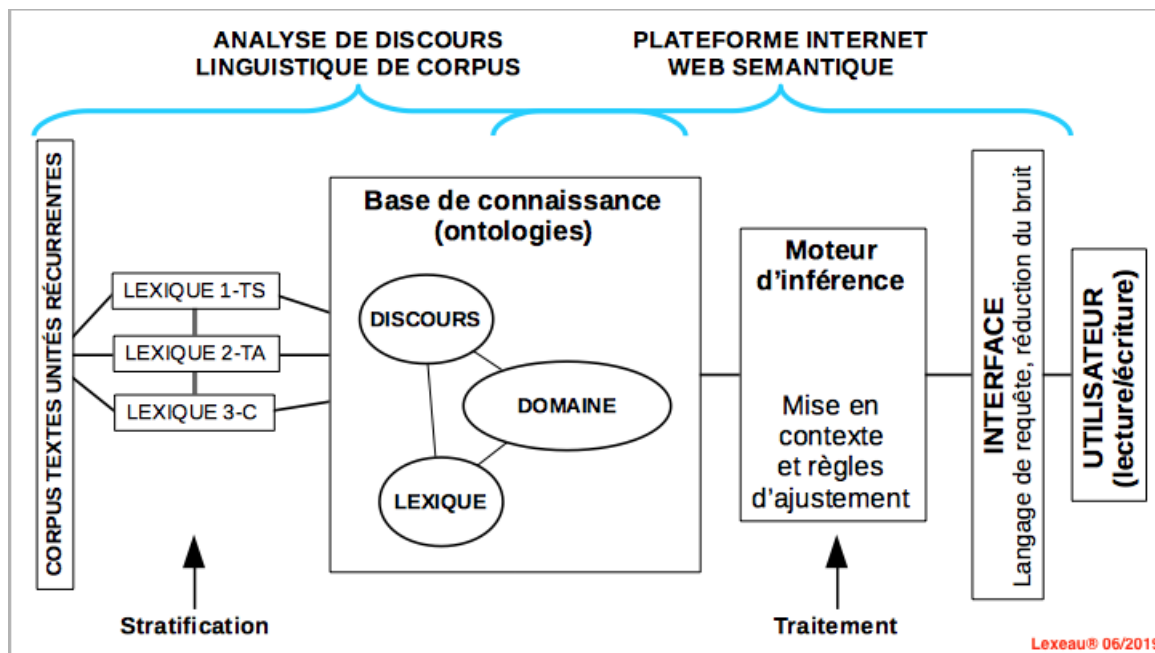


FIGURE 1 – Projet de lexique de l’eau stratifié : schéma du dispositif d’ensemble

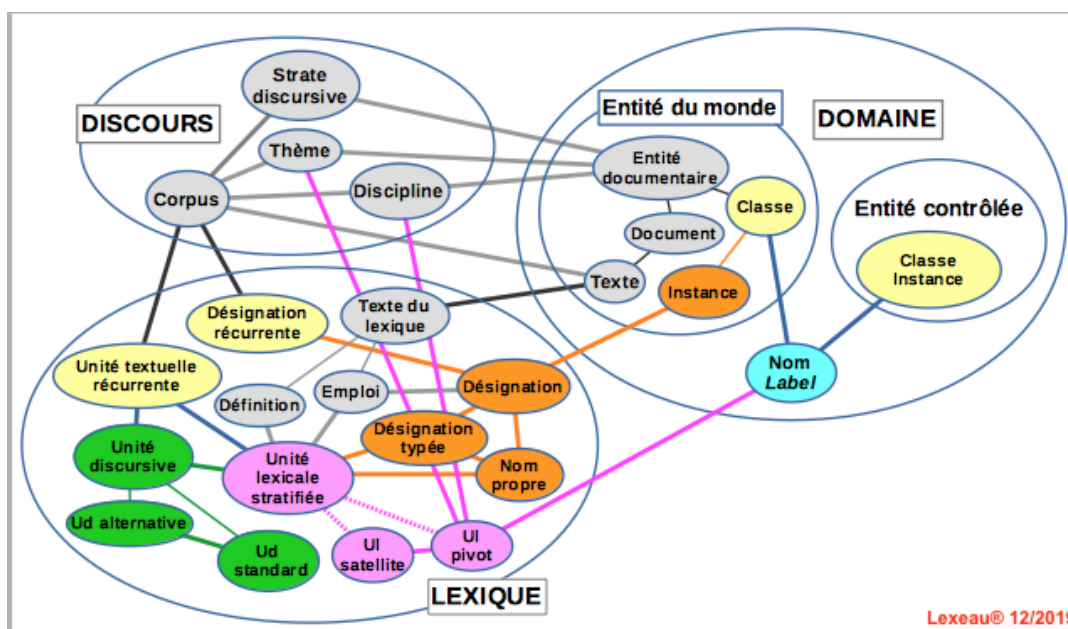


FIGURE 2 – Projet de lexique de l’eau stratifié : partition de la base de connaissance