

# Le dictionnaire de l'Académie de l'Eau

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

Jean-Louis JANIN\*  
Henri PORTINE†

Assemblée générale de l'Académie du 14 octobre 2019

---

\*ICPC honoraire, Docteur en linguistique (contact : [chevalierjlj@free.fr](mailto:chevalierjlj@free.fr))

†Professeur émérite de linguistique - Université Bordeaux Montaigne

### 1 Introduction

Tous les domaines se heurtent à la question du langage : partager ou non les mêmes mots mais en sachant opérer les applications (plus généralement surjectives que bijectives) nécessaires lorsque les mots divergent ; mettre le même sens sous les mots partagés ; structurer les énoncés de façon compatible. 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, un lexique scientifique et des lexiques de la quotidienneté (agents du terrain, clients, journalistes et autres médiateurs) suppose ainsi des formes d'appariement et d'ajustement parfois complexes ; c'est pourquoi les lexiques bijectifs échouent. Il faut donc être plus ambitieux, recueillir les usages effectifs, les mettre en regard avec des outils qui permettent des évaluations fines. C'est ce que vise le projet de Jean-Louis Janin en composant une structure qui articule des lexiques et — aussi parce qu'il faut bien des pivots pour assurer les interrelations — une ontologie du domaine de connaissance de l'eau. Ce faisant, ce projet apporte, en même temps, une contribution à l'élaboration de lexiques spécialisés ou non. Le projet a été élaboré en français et en anglais, sans prééminence d'une langue sur l'autre ni terminologie imposée, ce qui revient à accepter des concepts et des entrées intraduisibles et des unités discursives alternatives. Cette note présente le projet de l'Académie de l'Eau et un résumé de la thèse<sup>1</sup> consacrée au projet. La thèse a permis d'élaborer une maquette du prototype à réaliser en open source pour tester collectivement une plateforme de consultation du dictionnaire bilingue. Les figures 1 et 2 extraites de la thèse présentent le dispositif d'ensemble du dictionnaire et la partition de sa base de connaissance.

### 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

---

1. « Élaboration d'un lexique de l'eau stratifié en fonction de l'auditoire : du concept aux lexèmes »

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 « pivot » qui peuvent être associées à des unités dérivées dans d'autres strates discursives. Par exemple, l'unité lexicale « pivot » *Prélèvement d'eau* est définie dans la strate technico-scientifique comme un mouvement d'eau situé à l'amont du *Flux d'eau anthropique* (autre unité lexicale « pivot ») qu'il génère. Elle a deux unités lexicales « satellites » *Prélèvement d'eau au sens de la police de l'eau* et *Prélèvement d'eau soumis à redevance* dans la strate technico-administrative et l'unité lexicale « satellite » *Prélèvement d'eau*, entendue comme un volume d'eau, dans la strate discursive « courante ». La consolidation et la mise à jour du lexique et de l'ontologie du domaine passent par l'analyse des textes qui fait émerger des unités linguistiques récurrentes et de nouveaux concepts.

Les tests de la maquette du projet effectués au cours de la rédaction de la thèse permettent de passer à la réalisation d'un prototype en faisant appel à un prestataire rémunéré (informaticiens) avec un groupe de projet et une équipe de réalisation (cahier des charges et tests du 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 lemmes<sup>2</sup> identiques ou proches. Les propriétés des classes de l'ontologie génèrent dans le lexique les relations lexicales classiques du type *Aléa* <est l'hyperonyme de> *Crue* et *Crue* <est l'hyponyme de> *Aléa*, mais aussi d'un nouveau type, par exemple *Crue* <déclenche> *Inondation de rivière*, ce qui fait du réseau des entrées du lexique stratifié un réseau « cognitif » avec son métalangage et ses inférences potentielles. Les classes d'entités typées de l'ontologie du domaine permettent d'associer un dictionnaire des noms propres aux noms de classe transposés dans le lexique (la *Garonne*, en tant que nom propre de ce *Cours d'eau*).

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 la question des prélèvements d'eau et sur celle des inondations, après modélisation des sous-domaines correspondants dans l'ontologie. La discussion, puis l'abandon, d'une quatrième strate discursive pour rendre compte des discours artistiques, religieux et idéologiques sur l'eau 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.

---

2. 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.

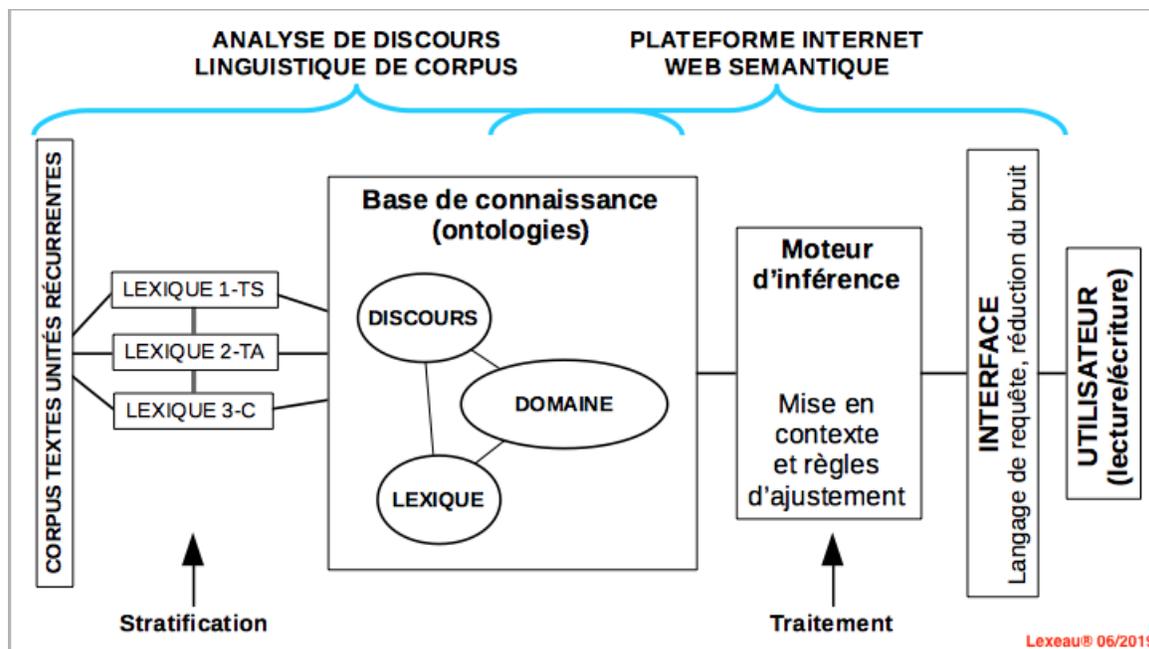


FIGURE 1 – Projet de lexique de l'eau stratifié : schéma du dispositif d'ensemble

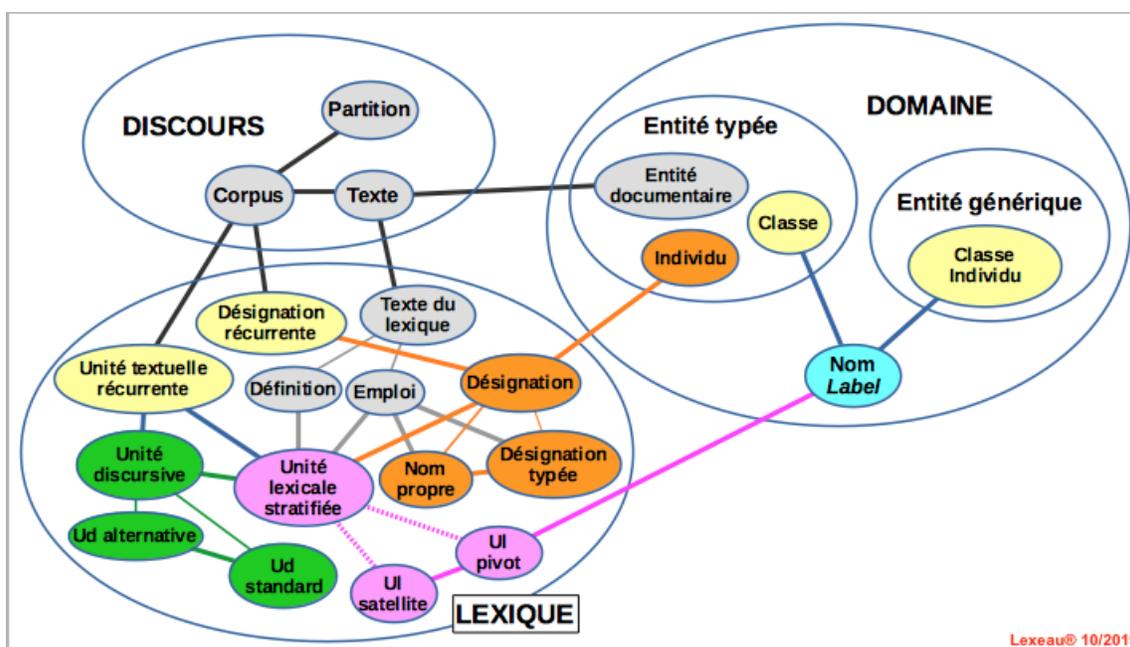


FIGURE 2 – Projet de lexique de l'eau stratifié : partition de la base de connaissance