

Legivoc: connecter les droits dans un monde en mouvement

Hughes-Jehan Vibert*, Pierre Jouvelot°, Benoît Pin°

* *Docteur en droit, Chef de projet TIC, Ministère de la Justice¹, France*

° *Chercheurs, MINES ParisTech², France*

Résumé. Sur l'internet, l'information juridique disponible est essentiellement une somme de droits nationaux. Même dans un monde en mouvement, même en acceptant une approche globalisée, le droit est d'abord une affaire de cultures (de cultures nationales le plus souvent) et les concepts juridiques n'ont de sens qu'accompagné de l'environnement juridique auquel ils sont rattachés. Face à ce constat, Legivoc vise à être une interface sémantique entre le sujet de droit d'un État et les différents espaces d'informations juridiques qu'il sera amené à utiliser. Ce projet consistera d'abord à mettre en place un serveur de vocabulaires juridique multilingue de chacun des États membres de l'Union européenne.

Mots clés: droit; langue; internet juridique; informatique juridique; thesaurus; vocabulaires; alignement; États membres de l'Union européenne

L'objectif de Legivoc est de créer une base de données terminologiques dont l'utilisation doit faciliter la compréhension des droits des États membres de l'Union européenne entre eux et de construire un système terminologique interopérable utilisable, notamment, dans les principaux projets d'informatique juridique relatif à l'accès au droit des États membres ainsi qu'à l'appui des échanges d'informations juridiques émises entre les réseaux européens de coopération juridique ou judiciaire. Dans cet environnement que l'internet appréhende mondialement, le projet offrira une forme de rationalité dans le traitement des informations juridiques permettant *in fine* de faciliter la recherche d'information et la compréhension des droits étrangers dans un ensemble européen structuré mais où les diversités juridiques s'expriment

¹ <http://www.justice.gouv.fr>.

² <http://www.mines-paristech.eu>.

indépendamment des diversités linguistiques. Legivoc est un projet initié par le ministère de la Justice français, avec le soutien des ministères de la Justice de la République Tchèque, d'Espagne, d'Italie, de Finlande et du Luxembourg, ainsi que de la Commission européenne.

1. Pourquoi et pour qui connecter les droits?

Alors que l'internet a considérablement accéléré les échanges d'informations dans un environnement international et multilingue, les recherches d'informations juridiques d'un État membre de l'Union européenne à l'autre ne peuvent être encore satisfaisantes du fait d'un obstacle dépassant celui de la barrière linguistique.

En effet, le droit est nécessairement celui d'une culture et les termes juridiques ne traduisent pas toujours la même réalité entre, par exemple, le Luxembourg et la France, entre l'Autriche et l'Allemagne ou même, entre l'Écosse et le reste du Royaume-Uni. Le thésaurus multilingue *eurovoc*³ constitue un outil précieux pour l'accès à l'information, y compris juridique, de l'Union européenne, il n'est toutefois qu'accessoirement juridique sans être adapté à la diversité des droits des États membres. De son côté l'IATE⁴, la base de données terminologique multilingue de l'Union européenne, offre un outil tout aussi indispensable et précieux, constitué d'une somme de sources internationales qui n'est toutefois pas toujours consacrée aux droits des États membres.

La pertinence de la représentation sémantique juridique d'un État ou d'une organisation n'a de sens que dans ce cadre délimité. Ainsi, si le domaine « droit » d'eurovoc est l'outil idéal pour apprécier l'information juridique de l'Union européenne, il ne l'est plus tout à fait pour celle de ses États membres où s'expriment des traditions romano-germaniques, mixtes et de Common Law avec des spécificités nationales. Il y a, par définition, une intelligence

³ eurovoc.europa.eu.

⁴ iate.europa.eu.

collective des sociétés⁵, c'est l'articulation sémantique entre ces intelligences que Legivoc se propose d'apprécier.

1.1 LES ELEMENTS A CONNECTER

1.1.1 *Les droits*

Au niveau européen, où la nécessité de pouvoir apprécier les droits nationaux est consubstantielle à l'édification institutionnelle de l'Union, comme au niveau plus général de la recherche en droit comparé, la mise en évidence de relations sémantiques entre concepts juridiques exprimés dans des langues différentes et des droits différents est un idéal complexe à formaliser. Ce travail préalable a déjà été pris en compte par les institutions européennes et d'ailleurs, en pratique, il entre au cas par cas dans les études préalables aidant à la négociation des textes communautaires.

C'est pour répondre à ce type de besoins que la Direction générale de la Traduction (DGTrad) de la Cour de justice de l'Union européenne a mis en œuvre un projet pilote (le V.J.M ou *vocabulaire juridique multilingue*) permettant de disposer d'un outil documentaire multilingue pour indexer et interroger les fonds documentaires des bases de données nationales juridiques à partir de n'importe quelle langue et de manière uniforme. Ce travail est immense, considérable par les diversités juridiques qu'il met en évidence et les coûts qu'il engendre.

Depuis le 1^{er} juillet 2013, Il y a 28 États membres dans l'Union européenne et 24 langues officielles. La DGtrad de la CJUE a estimé en 2010 qu'il fallait 34 jurilinguistes, travaillant 20 semaines pour traiter 400 notions⁶. À titre de comparaison, le thésaurus belge UTU utilisé par le pouvoir judiciaire belge,

⁵ Fazy, James, *De l'Intelligence collective des sociétés, cours de législation constitutionnelle, donné par M. James Fazy*, Genève : 1873, Impr. de Vve Blanchard, 450 p.

⁶ Cour de justice de l'Union européenne | Direction générale de la Traduction, *Vocabulaire juridique multilingue comparé Bilan sur l'étude de faisabilité*, par Caroline Reichling, 30 juin 2010, en ligne: http://curia.europa.eu/trad/bilan_vocabulaire_juridique_multilingue_compare.pdf.

contient 8000 termes. Par un calcul, absurde, un traitement de la DGtrad à partir de ces seuls termes, propres au droit belge (qui s'exprime en 3 langues officielles) demanderait plus de 8 ans. Pour poursuivre dans cette voie, si l'on met la lumière sur une initiative lithuanienne regroupant 40000 termes de droit et de sciences politiques lithuanien traduits en français⁷, ce calcul devient un doux rêve sur plus de 40 années démultipliées par une complexification inévitable lorsqu'il s'agit de réaliser une corrélation sémantique entre une trentaine d'États (puisque l'Union européenne à vocation à s'étendre).

Legivoc vise à accompagner cette démarche complexe et ambitieuse comme à être utilisé, à son tour, pour d'autres développements ultérieurs. En effet, nous considérons que le travail déjà réalisé par la DGTrad de la CJUE constitue un idéal. Car un traitement humain, par des jurilinguistes, des concepts juridiques nationaux entre eux, en vue de leur mise en relation constitue l'approche la plus exacte, la moins sujette aux pertes de sens que provoquerait un traitement automatique et/ou reposant sur une simple traduction littérale, et les quelques thématiques déjà traitées par la DGTrad ne devront plus être à refaire, constituant au contraire un socle sur lequel nous pouvons nous fier indépendamment des inévitables erreurs ou incomplétudes pouvant se glisser ça et là.

Mais Legivoc repose d'abord sur les contributions des États membres de l'Union européenne et, le cas échéant, sur les traitements visant à réaliser des thésaurus juridiques nationaux. L'approche de Legivoc est presque simpliste consistant d'abord à regrouper les droits nationaux, selon leur organisation propre, et selon leur(s) langue(s) officielle(s). Après une demande faite aux États membres de l'Union européenne, c'est près d'une quinzaine de vocabulaires juridiques nationaux qui pourront prochainement être mis à disposition aux fins de préparer la mise en relation conceptuelle mais également, à la condition que les États l'autorisent, aux fins de tous usages, sur le plan national comme pour toute approche en droit comparé.

⁷ *Lietuvių - Prancūzų- teisės ir politikos žodynas*, par Markas Paura en ligne, <http://www.zodynastp.lt>.

1.1.2 Les langues ?

La notion de *contrat écrit* s'exprime en lithuanien (et en Lituanie) par *sudaryti rašytinę sutartį*, la compréhension d'une telle notion passe naturellement par la langue mais il s'agit de ne pas s'arrêter à une simple traduction littérale. Qu'est-ce que le *contrat* en droit lithuanien ? Qu'est-ce que l'écrit et, implicitement, la *preuve* ? Issu d'un autre État les auteurs ont toutefois des présupposés pour répondre à ces questions mais, fondamentalement, nous ne pouvons faire d'affirmation sans risques d'erreurs. Il s'agit en effet d'apprécier pour chaque concept juridique un ensemble de notions liées réclamant une certaine connaissance du système juridique considéré.

Une traduction littérale est une facilité qui provoque de sérieux risques de compréhension et des erreurs d'interprétation. Le jurilinguiste Guy Mazet qui a résumé 30 ans de recherches dans sa thèse⁸ nous donne ainsi l'exemple des droits italiens et français autour du terme *amministratore* et *administrateur*, qui semblent bien équivalents mais où le premier s'emploie [en Italie] pour l'ensemble des dirigeants quelque soit la forme sociale et où le second ne s'applique, en France, qu'aux seules *sociétés anonymes*. Mais qu'est-ce qu'une *société anonyme* hors de France ? S'il existe, qu'en est-il de l'*administrateur* en droit belge ou en droit luxembourgeois alors qu'il n'y a pourtant pas d'obstacle linguistique dans ces États où le droit s'exprime en français, seul ou avec d'autres langues officielles ? Les outils de traduction automatique sont-ils à même de considérer les subtilités qui émaillent de l'ensemble du corpus juridique de chaque État, au nom de sa culture, de son histoire ou de sa politique, forcément originale ? Certainement pas ou, tout au moins, pas sans un travail préalable que Legivoc se propose justement de représenter.

Dans sa thèse de doctorat, Guy Mazet estime que « *la conception d'un instrument bilingue d'accès à des bases de données juridiques ne se réduit pas à un problème de simple traduction. Les rapports d'équivalence linguistiques ne rendent pas compte, à eux seuls, des modalités de la rencontre des systèmes*

⁸ Mazet, Guy, *Jurilinguistique et informatique juridique*, 2001, IRETIJ, Université Montpellier I.

juridiques. Il convient de les conserver lorsqu'ils se présentent opportunément mais il faut rechercher le niveau de relations intrinsèques existant entre les deux ordres juridiques ; ce niveau est celui de l'équivalence fonctionnelle »⁹. Avec plus de 28 systèmes juridiques (sans compter les spécificités nationales) où 24 langues officielles s'expriment (sans compter les langues nationales qui ne sont pas langues officielles de l'Union européenne) Legivoc consacre de façon massive, mais tout aussi strictement, cette analyse de Mazet.

1.2 IN VARIETATE CONCORDIA¹⁰

Des langues, des droits, des liens entre les droits, tels sont les éléments à notre disposition et tel pourrait être résumé le projet Legivoc. À la recherche d'équivalence fonctionnelle nous considérons que la langue n'a que peu d'importance dans ce projet. Si la traduction intéresse le projet, elle n'est qu'un outil pouvant aider ou au contraire nuire à la compréhension juridique des termes présentés, selon sa qualité et les connaissances préalables qu'elle demande. Sur Legivoc, comme dans le droit des États membres de l'Union européenne, des seules langues ayant du sens seront les langues officielles des États considérés.

Cette approche nous donne une grande souplesse. Elle permettra presque paradoxalement d'alimenter ultérieurement, si le besoin s'en fait sentir, la base de données du projet d'un ensemble de traductions potentiellement infini. En effet, le droit français exprimé en anglais, en allemand ou en espagnol a autant de légitimité que s'il était exprimé dans des langues tierces à celles de l'Union européenne, de l'arabe au russe en passant par le chinois, le turc, ou l'hindou. Puisque seule la langue officielle du droit de France a du sens, toutes les traductions n'ont qu'une valeur indicative et, encore une fois, seule la mise en

⁹ *Ivi*, p. 18.

¹⁰ *Unie dans la diversité* est la devise de l'Union européenne, «cette devise caractérise la manière dont les Européens se sont rassemblés, en créant l'Union européenne, pour œuvrer en faveur de la paix et de la prospérité, s'enrichissant ainsi des diverses cultures, traditions et langues du continent»,

http://europa.eu/about-eu/basic-information/symbols/motto/index_fr.htm.

relation fonctionnelle des concepts juridiques nationaux est susceptible d'offrir une réelle traduction juridique en fonction de l'État considéré aux fins de comparaison mais également du destinataire de la traduction.

En fonction de ces éléments, notre approche sera pratique et rationnelle. Dans un premier temps, tous les vocabulaires juridiques nationaux mis à disposition sur le projet Legivoc seront disponibles dans l'ensemble des langues officielles de l'Union européenne. Le vocabulaire juridique national mis à disposition dans une ou plusieurs langues officielles par un État sera la seule référence et ses traductions seront mises à disposition comme simples indications. Ces suggestions seront réalisées automatiquement par les outils de la Direction Générale de la traduction de la Commission européenne (MT@EC, le service de traduction automatique pour les langues officielles de l'Union européenne¹¹). Sur ces insertions automatiques, les utilisateurs auront la possibilité de fournir d'autres suggestions.

¹¹ http://ec.europa.eu/isa/actions/02-interoperability-architecture/2-8action_en.htm.

Section *droit belge* de legivoc (exemple)

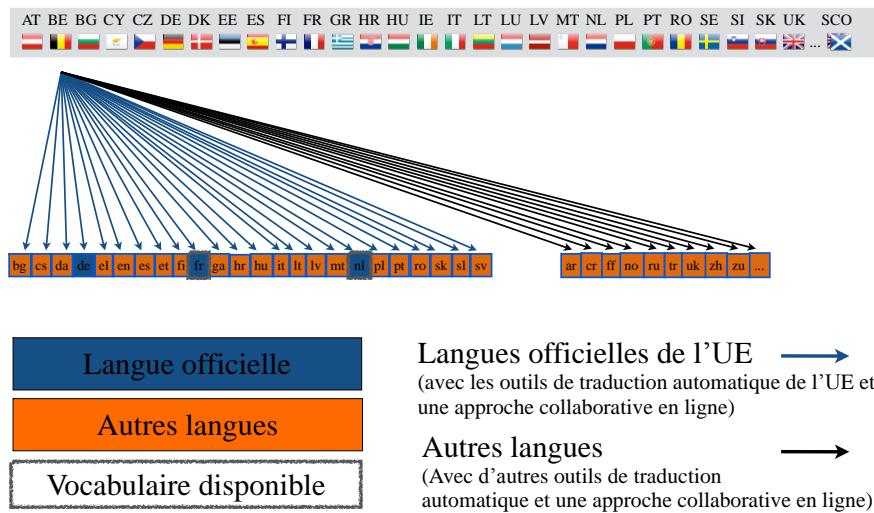


Fig. 1: l'approche de Legivoc, par État, exemple de la Belgique dont nous disposons du vocabulaire juridique dans deux de ses trois langues officielles

Dans l'illustration ci-dessous nous voyons mis en avant les informations à notre disposition, le thesaurus juridique du droit belge UTU, disponible dans 2 des 3 langues officielles de Belgique. Dans cette illustration, la notion de *demande additionnelle* ou *aanvullende eis* (*additional claim* ou, plutôt, *additional demand* en droit belge anglicisé) correspond à un unique concept propre au droit belge et matérialisé par le numéro 20. Naturellement, ce concept numéro 20 pourra être traduit dans toutes les langues possibles, ces traductions suivront l'organisation hiérarchique du thesaurus belge, 1 -> 2 -> 11 -> 13 -> 20. Ici le droit réside dans les adresses et leurs références, pas dans les mots.



| DROIT JUDICIAIRE | 1 | GERECHTELIJK RECHT |
|---|------|---|
| DROIT JUDICIAIRE - PRINCIPES GÉNÉRAUX | 2 | GERECHTELIJK RECHT- ALGEMENE BEGINSSELEN |
| Principes (droit judiciaire - principes généraux) | 3 | Beginselen (gerechtelijk recht - algemene beginselen) |
| Généralités | 4 | Algemeen |
| Droit commun de la procédure | 5 | Gemeen procesrecht |
| Droit transitoire | 6 | Overgangsrecht |
| Déni de justice | 7 | Rechtswegering |
| Prononciation par voie de disposition générale et réglementaire | 8 | Uitspraak algemene regel |
| Lois interprétatives | 9 | Uitleggingswetten |
| Compétence | 10 | Bevoegdheid |
| Notions | 11 | Begrippen |
| Généralités | 12 | Algemeen |
| Demande | 13 | Vordering (eis) |
| Généralités | 14 | Algemeen |
| Demande | 15 | Eis |
| Demande incidente | 16 | Tusseneis |
| Demande reconventionnelle | 17 | Tegeneis |
| Demande en intervention | 18 | Tussenkomst |
| Demande nouvelle | 19 | Nieuwe eis |
| Demande additionnelle | 20 | Aanvullende eis |
| Jugement - arrêt | 21 | Vonnis - arrest |
| Généralités | 22 | Algemeen |
| ... | | |
| Politique étrangère et de sécurité commune | 8381 | Gemeenschappelijk buitenlands en veiligheidsbeleid |
| Énergie - Droit européen | 8382 | Energie - Europees recht |
| Protection civile - Droit européen | 8383 | Civiele bescherming - Europees recht |

Fig. 2: extrait du thesaurus belge UTU, dans 2 des 3 langues officielles

Dans cette partie, pourtant fondamentale, le projet Legivoc ne crée donc rien, se contentant de représenter ce que les États membres de l'Union européenne à travers leurs ministères, ordres professionnels universitaires ou bibliothèques juridiques fourniront au projet. Nous initiions une simple approche afin de permettre des rapprochements. Au premier septembre 2013, nous disposons d'une quinzaine de vocabulaires juridiques nationaux sur 28 États membres, disponibles le plus souvent sous un format facilement exploitable, parfois structurés et organisés, parfois simple liste de mots, parfois traduits dans d'autres langues, parfois exhaustif, parfois très incomplets.

Par cette dimension multinationale, mult droit et multilingue, multiformat et multiapproche où les cultures juridiques sont présentés dans leurs espaces propres et représentés sur d'autres sphères linguistiques, Legivoc est un simple révélateur de différences qu'il s'agira de conserver y compris dans l'effort de rationalité que constitue la phase expérimentale de l'alignement sémantique des concepts juridiques. C'est en fonction des usages pressentis et de ceux qu'il reste à dessiner que nous tenons à conserver une structure la plus souple possible dans l'organisation et la mise à disposition des concepts juridiques nationaux.

2. Comment les droits se connectent sur Legivoc?

Nous disposons actuellement d'une quinzaine de vocabulaires nationaux dont il s'agira de conserver la logique propre tout en assurant une mise à disposition sous un format unique. La base de données donnera accès aux vocabulaires qui nous ont été transmis et offrira la possibilité ultérieure de mettre à disposition les vocabulaires dans les traductions qui auront été ajoutés selon les besoins. Cette dimension du projet le rend sans doute déjà intéressant, elle constitue un préalable indispensable à d'autres usages terminologiques et qu'il ne sera plus nécessaire de mettre en place, d'autres pouvant se charger d'améliorer le traitement de ces vocabulaires (en fonction des licences d'utilisation propres à chaque État).

Nous l'avons vu, Legivoc disposera du travail réalisé par la Direction générale de la Traduction (DGTrad) de la Cour de justice de l'Union européenne. Ce travail consiste notamment en une recherche d'équivalent conceptuels et donc une mise en relation des concepts juridiques nationaux entre eux. Nécessitant l'emploi d'une trentaine de jurilinguistes sur plusieurs années, ce travail est dispendieux tout en étant qualitativement idéal. L'apport humain est en effet ici indispensable. Nous proposons de conserver cette approche humaine par l'utilisation des interactions entre différents espaces d'échanges. Deux projets tiers sur lesquels nous avons une maîtrise technique serviront de support au projet Legivoc qui sera préalablement mis à disposition de ces projets : Legicoop, le site du réseau de coopération législative des ministères de la

Justice de l'Union européenne (RCLUE) et Babellex, place de marché virtuelle permettant de réunir les offres et demandes de traducteurs et interprètes juridiques et de regrouper le travail réalisé dans une bibliothèque de traduction.

2.1 L'ALIGNEMENT SUR DES PROJETS CONTROLES

2.1.1 *Legivoc et le réseau de coopération législative des ministères de la Justice de l'Union européenne*

Le site du RCLUE suit une triple dimension à la fois culturelle, technique mais également humaine puisque le site d'un réseau d'échange d'informations se devait d'être participatif. Au moment de la conception du site, le principal enjeu était de ne pas limiter la dynamique installée lors des 3 premières années d'existence du RCLUE, mais au contraire de favoriser son développement.

Depuis le 1^{er} juillet 2013, 24 langues officielles caractérisent l'Union européenne et 28 droits (et plus en prenant en compte les régions, lands et nations) s'y expriment. Une devise européenne, *in varietate concordia*, résume ce contexte qu'il fallait suivre également. L'aspect technique lié à l'internet et à cette virtualisation du réseau ne devait pas masquer la dimension culturelle dans laquelle des individus seront à même de s'exprimer, d'agir et d'interagir. 29 correspondants nationaux (le Royaume-Uni dispose de 2 correspondants Angleterre/Pays de Galles et Écosse) ont été désignés par leurs ministères respectifs au fil du temps. L'outil mis en place vise à lier ces diversités à des fins de connaissances réciproques autour de questions de droit communes.

Au service de ses utilisateurs, le site a été construit autour du plus petit dénominateur commun quant à la maîtrise de l'informatique. En effet, il apparaissait pertinent de ne pas modifier le mode de transmission des échanges en maintenant quelques éléments du fonctionnement de la liste de diffusion servant originellement au RCLUE. Le courriel a donc été le mode de diffusion privilégié pour les échanges: à chaque question ou réponse déposée sur le site Legicoop, un courriel est transmis aux correspondants contenant un lien visant à découvrir la question et/ou pouvoir y répondre directement sur le site.

Legicoop, le site du RCLUE, permet donc d'associer dans le même mouvement et directement, les échanges et la communication entre les correspondants et l'alimentation du site.

Le site a donc été pensé de façon à être simple dans un environnement complexe, à la fois multilingue et multi-droits et même multi-usages puisqu'il est adapté aux ordinateurs mais aussi aux téléphones mobiles et tablettes (utilisées dans certains ministères).

legicoop

Communication language (same as before)

Title of the request

Descriptor

Request

(Auto)Reply for the Italian law

Attached documents (translated or not)

Attached links

National language

Titolo della richiesta

Descrittore

Richiesta

Rispondi per la legge italiana

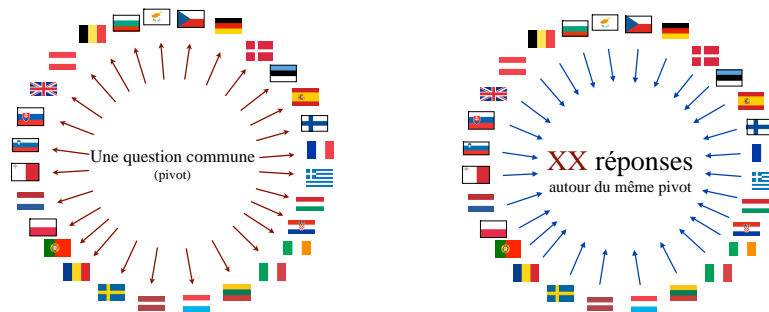
legivoc

required

Fig. 3. Site web du RCLUE vue schématique

Les correspondants nationaux s'expriment sur le site du RCLUE dans une langue de communication comprise par le ou les pays requis, souvent en anglais mais ils doivent ajouter des informations dans leurs propres langues nationales (le titre de la question et quelques mots clés, dans cet exemple donné pour le correspondant de l'Italie). Ce sont ces informations qui seront issues de Legivoc. Le correspondant national disposera donc d'un vocabulaire national de référence qui aidera à améliorer qualitativement l'information transmise.

Très simplement, les échanges tournent autour d'une question commune et à une question posée par un État, les autres États donnent la solution propre à leur système juridique. Le schéma ci dessous illustre cette simplicité et c'est autour de cette question juridique commune que se réalise, questions après questions une mise en relation des concepts juridique.



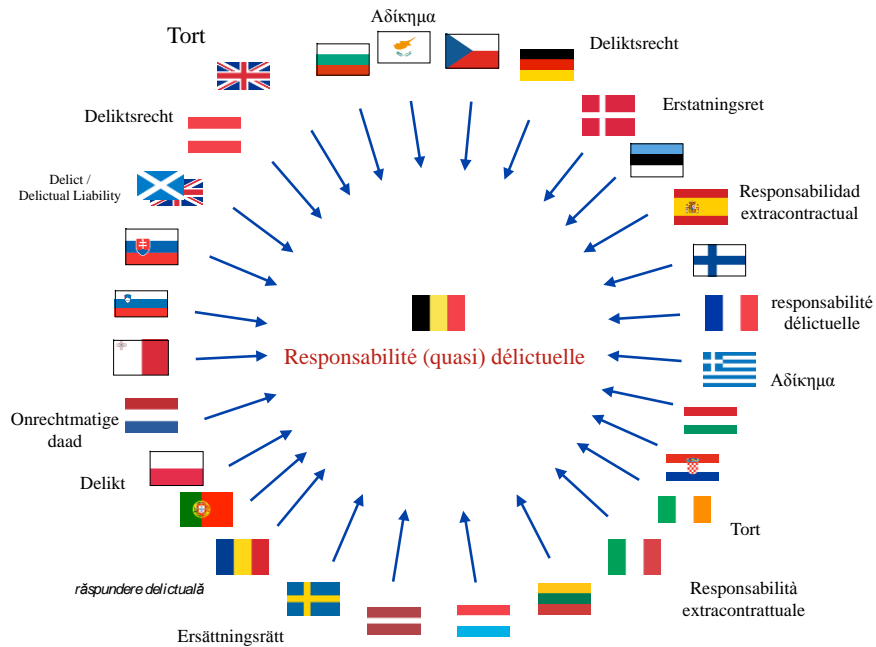


Fig. 4. Le pivot de connexion sera une question juridique, ce ne sera pas une langue (exemple juridiquement approximatif)

Naturellement, la pertinence des mises en relation de concepts juridique nationaux prendra du temps et seule la multiplication des échanges permettra d'espérer cette pertinence qui restera soumise à certaines approximations. C'est pour cela que cet outil, Legicoop, financé par l'Union européenne et 5 ministères de la Justice d'États membres sera offert aux autres réseaux d'échange d'information qui en feront la demande. L'utilisation du système Legicoop entraînant celle de Legivoc. À son tour, le serveur de Legivoc récupèrera les connexions réalisées, autour des seuls mots clés, dans le cadre des échanges et selon des modalités qui seront développées plus loin.

2.1.2 *Legivoc et Babellex*

Babellex est un projet en développement qui est monté conjointement avec le projet Legivoc. Babellex est soutenu par la France, l'Estonie, l'Autriche, l'Allemagne et la Commission européenne. Il est coordonné par le ministère de la Justice français. Ce projet prévoit la constitution d'un espace où les traducteurs juridiques pourront être sollicités dans le cadre de toute procédure judiciaire nécessitant une traduction ou un interprétariat¹². Le projet visera également à constituer une bibliothèque de traduction juridique des textes traduits et notamment des actes judiciaires. C'est sur cette seconde phase du projet que Legivoc sera enrichi de par son utilisation pour l'aide à la traduction des concepts juridiques¹³ par les traducteurs.

Nous reproduirons donc une approche participative (crowdsourcing) et donc humaine de mise en relation de concepts nationaux mais, cette fois-ci, dans une simple approche de traduction. Initié dans un espace européen le projet concerne potentiellement toutes les langues écrites. Par cette approche, quelques traductions de concepts juridiques d'États membres de l'Union européenne pourront être rendus disponibles sous une forme non plus automatique mais humainement contrôlée.

Ces espaces de mise en relation de concepts juridiques nationaux vis-à-vis d'autres ou aux fins de traduction sont précieux. Il s'agit du travail d'experts, juristes nationaux les plus à même de qualifier finement leur propre droit ou traducteur juridique professionnels. Ces expertises sont le cœur de ces mises en relation et bénéficieront d'un taux de crédibilité très fort. Et, puisque nous parlons chiffres en pensant aux statistiques, nous pouvons également envisager l'enrichissement de Legivoc par un public bien plus large. À cette fin, un usage didactique de Legivoc sera sans doute offert, permettant à chacun de proposer des mises en relation (de langues et de droit), à charge pour nous d'apprécier la

¹² Voir la *Directive 2010/64/UE du Parlement européen et du Conseil du 20 octobre 2010 relative au droit à l'interprétation et à la traduction dans le cadre des procédures pénales*, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32010L0064:FR:NOT>.

¹³ Il appartiendra aux États de préciser les droits d'utilisation et de réutilisation des traductions, des textes comme des concepts.

pertinence de ces contributions libres. Ici encore l'API Legivoc sera la porte d'entrée vers ces enrichissements et qu'il s'agira donc de détailler.

2.2 CONNEXION VIA D'AUTRES SYSTEMES: LEWIS (LEGIVOC WEB INFORMATION SYSTEM)

Gérer les vocabulaires textuels légaux de plusieurs dizaines de pays ne peut se faire que via un système d'information puissant et néanmoins flexible, que nous appelons pour le moment LeWIS (*Legivoc Web Information System*). Un de ses objectifs principaux est d'éviter d'imposer des modifications aux procédures existantes d'échange de données des divers partenaires, tout en fournissant une voie d'évolution nouvelle et générale à tous ceux qui souhaitent adopter ce que nous considérons comme "meilleures pratiques".

L'architecture cible de LeWIS est donnée dans la figure 5. Elle est structurée autour d'une base de données principale qui stocke les thésaurus et vocabulaires législatifs fournis par les nations participantes. Etant donné le manque actuel de spécifications normatives concernant les données législatives, les divers intervenants juridiques utilisent différents formats pour le stockage de leurs informations et adoptent des définitions sémantiques différentes pour celles-ci. Pour s'adapter de manière fluide à cet état de fait, nous envisageons de développer des traducteurs pour exprimer les divers formats existants dans notre propre format interne Legivoc (suffixe .lgv). La définition du langage spécifique (DSL) LGV, que nous implémenterons probablement comme dérivation XML du *Simple Knowledge Organization System (SKOS) Core*¹⁴, est en cours d'élaboration. Une fois cette définition de langage finalisée, notre objectif est que LeWIS puisse accepter tout type de fichier d'entrée encodé en XML (.xml), Excel (.xls), texte (.txt) et, in fine, Legivoc (.lgv), si les partenaires décident d'adopter ce format. Étant donné les mises à jour quasi permanentes des vocabulaires légaux dans l'ensemble de l'Europe, automatiser ces processus de transcodage sera un accélérateur de productivité certain. Bien évidemment, le procédé inverse, permettant de traduire des fichiers .lgv en divers types de formats de sortie existants, sera

¹⁴ www.w3.org/2004/02/skos.

également disponible, pour permettre d'utiliser les outils préexistants.

Puisque que l'offre d'accès multimodaux aux vocabulaires est au cœur du projet Legivoc, nous envisageons de développer deux modes d'accès à cette base de données : une interface Web et une API (*Application Programming Interface*). Ces deux interfaces seront accessibles via un serveur, ce qui assurera que LeWIS sera accessible par un large choix de clients, des ordinateurs de bureau aux téléphones mobiles. La première interface utilise un navigateur. Une fois en ligne, le site Legivoc.eu permettra aux clients authentifiés d'accéder aux concepts légaux, législatifs et juridiques fournis par les partenaires de chaque pays, en parallèle avec leurs traductions dans toutes les langues officielles de l'Union européenne. Ces pages (un concept par page HTML) contiendront également des informations supplémentaires, quand elles sont disponibles, concernant des concepts voisins, des synonymes, des homonymes... et leurs traductions ; la représentation de ces relations sera fondée sur les capacités de liens proposées par SKOS. Pour accélérer le développement de cette interface utilisateur et constituer aussi rapidement que possible une communauté autour d'elle, nous envisageons la possibilité d'utiliser un des systèmes de gestion de thésaurus existants.

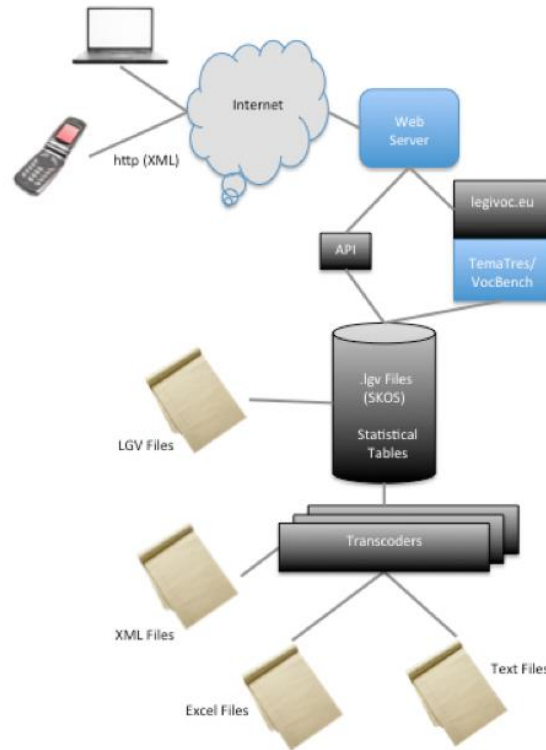


Fig. 5. Dans ce schéma d'architecture, les boîtes noires sont propres à LeWIS

L'un d'eux est TemaTres¹⁵, qui est fondé sur le langage PHP et le système de base de données (DBMS) MySQL. Ce même DBMS est au centre également de VocBench¹⁶, un système conçu en Java. Ces deux systèmes ont des avantages et des inconvénients propres : par exemple, TemaTres est un système ouvert (*open-source*) et a l'avantage d'être suivi par des communautés actives de développeurs et d'utilisateurs, alors que VocBench offre l'avantage d'utiliser

¹⁵ <http://www.vocabularyserver.com>.

¹⁶ <http://aims.fao.org/tools/vocbench-2>.

un langage d'implémentation plus mature, à savoir Java. Qu'un des systèmes précédents soit choisi pour Legivoc ou qu'un système dédié soit déployé, nous nous assurerons que le même outil, à savoir SKOS, pourra être utilisé pour coder et stocker les fichiers .lgv.

La seconde interface permettra à des programmes tiers d'accéder aux données Legivoc via Internet en utilisant une interface de programmation (API) qui va être définie sous peu. Suivant le modèle de conception *Representational State Transfer* (RESTful)¹⁷, bien établi, elle permettra un accès direct à la base de données LeWIS, via divers modes d'interrogation et de mise à jour¹⁸. Même si la plupart des systèmes de gestion de thésaurus fournissent déjà des mécanismes d'accès via leurs propres APIs, nous envisageons de développer indépendamment notre API LeWIS. Une des raisons de ce choix est que, même si les aspects de gestion et de sécurité ne sont pas au cœur de notre projet pour le moment, LeWIS ambitionne d'offrir dès le début les moyens de déployer, à terme, via cette API, un service sécurisé d'accès aux données. PHP, utilisé dans TemaTres, n'est peut-être pas le meilleur outil pour atteindre cet objectif ; par ailleurs, TemaTres n'est pas un cadre très robuste, ce qui fait qu'étendre son ensemble de services n'est peut-être pas aussi aisé qu'on pourrait l'espérer. VocBench a plus d'atouts sur ce point particulier.

Un des objectifs clés de notre nouvelle API est de permettre à LeWIS de garder trace, sans gêner les utilisateurs, de leurs activités (et ce de manière détaillée, tout en restant anonyme) et de déduire de ces accès des informations sémantiquement riches. Un ensemble dédié de tables de bases de données au sein du DBMS MySQL sera constamment tenu à jour, à des fins d'analyse statistique, à la "Big Data". Un objectif à long terme de ce processus d'acquisition de données utilisateur est de déterminer des corrélations statistiques entre les différents modes d'accès et d'utilisation, ce qui devrait nous aider à établir des relations sémantiques entre les concepts législatifs, à la fois dans un pays donné et entre pays. Par exemple, l'ensemble des termes mentionnés par un utilisateur lors d'une session particulière peut suggérer que

¹⁷ <http://www.ibm.com/developerworks/webservices/library/ws-restful>.

¹⁸ Seuls des contributeurs identifiés auront le droit de modifier les données clés de LeWIS.

ceux-ci sont sémantiquement corrélés ; c'est encore plus vrai si l'information concernant le niveau de compétence des utilisateurs (par exemple juriste, néophyte ou traducteur officiel) est disponible.

Lors du développement de LeWIS, nous envisageons de suivre une approche inspirée de la méthode de recherche-action introduite par Kurt Lewin, en couplant fortement, de manière continue, les phases de conception, d'implémentation et d'expérimentation, avec l'aide des experts juristes. Pour le développement des processus de *workflow* évoqués ci-dessus, nous comptons donc nous appuyer fortement sur une plateforme dédiée (basée sur des solutions existantes comme TemaTres ou VocBench, ou développés en interne), des outils largement diffusés (SKOS) et des méthodes de programmation agiles fondées sur des langages dynamiques et riches tels que Python.

3. Conclusion

Legivoc est une méthode et un ensemble de solutions reposant essentiellement sur le travail collaboratif de juristes en action. C'est en effet à travers des situations juridiques concrètes auxquelles il s'agit de donner des solutions, que les droits se connecteront d'un système juridique à l'autre par l'utilisation d'outils en ligne dédiés à l'information juridique transnationale. Mais en mettant en avant pour Legivoc, au nom d'un certain rigorisme, cette dimension humaine, nous espérons également offrir les bases d'autres traitements, y compris automatiques et desquels pourront apparaître cette fois les homonymies juridiques, termes implicites, contre sens et autres difficultés sémantiques, à travers les droits, à travers les langues. Mais finalement, la simple mise en place d'un ensemble de vocabulaires juridiques nationaux sur un même espace constitue déjà une prise en compte des droits en mouvements. Ils pourront bientôt être observés, directement et simplement, sous un unique format SKOS et, autant que possible, librement.