



ISSN 1866-5268

ISSN en ligne 2261-2750

Des technologies du langage au service du citoyen

Thomas François

F.R.S.-FNRS, CENTAL, IL&C, UCL, Belgique
thomas.francois@uclouvain.be

Cédrick Fairon

CENTAL, IL&C, UCL, Belgique
cedrick.fairon@uclouvain.be

Reçu le 27-04-2018 / Évalué le 19-05-2018 / Accepté le 30-06-2018

Résumé

Les technologies du langage se sont rapidement développées ces 20 dernières années et sont aujourd'hui utilisées quotidiennement par le grand public dans des applications comme la correction orthographique, la traduction automatique, la recherche d'information, etc. Dans cet article, nous présentons deux projets qui ont eu recours aux technologies du langage pour venir en soutien d'actions de politique linguistique : Recto/Verso et AMesure. Nous défendons l'idée que ces technologies représentent une opportunité pour illustrer des actions de politique linguistique à très large échelle et dans certains cas, pour contribuer à réaliser celles-ci.

Mots-clés : technologies du langage, traitement automatique du langage, politique linguistique, lisibilité, nouvelle orthographe

Sprachtechnologie im Dienste der Allgemeinheit

Zusammenfassung

Das Gebiet der Sprachtechnologie hat in den letzten 20 Jahren eine rasante Entwicklung durchgemacht, was zur Folge hat, dass Applikationen mit automatischer Rechtschreibungskorrektur, Übersetzung und Informationssuche heute ein unumgänglicher Teil des täglichen Lebens der Bevölkerung sind. In diesem Artikel werden zwei Projekte vorgestellt, die Sprachtechnologie in den Dienst der Sprachpolitik stellen. Diese Technologie stellt nach Ansicht der Autoren dieses Artikels die Gelegenheit dar, nicht nur die sprachpolitischen Aktionen in weiterem Sinne darzustellen, sondern auch zu ihrer Umsetzung beizutragen.

Schlüsselwörter: Sprachtechnologie, automatische Sprachverarbeitung, Sprachpolitik, Lesbarkeit

Language Technologies in the service of citizens

Abstract

Language Technologies have developed dramatically over the past 20 years. They are now of daily use by the general public in numerous applications such as spell-checking, machine translation, information retrieval, etc. In this paper, we present two projects which have made use of language technologies to support actions inspired by linguistic policy. We argue that these technologies represent a real opportunity for illustrating initiatives in the domain of language policies, and in some cases for contributing to achieving their objectives.

Keywords: language technologies, natural language processing, linguistic policies, readability, spelling, new French orthography

Introduction

Les origines du *traitement automatique du langage* (TAL) se confondent avec celles de la traduction automatique que l'on peut situer aux États-Unis à la fin des années 1940. Vingt ans plus tard, la traduction automatique fut remise en question et de nombreux autres objectifs de recherche furent progressivement définis : aide à la traduction, recherche et extraction d'information, réponse automatique à des questions, etc. Depuis les années 2000, le secteur des technologies du langage tout entier connaît un développement à très vive allure, propulsé par les progrès de l'informatique (vitesse de processeur et capacité de stockage en progression constante, miniaturisation, etc.) et surtout par le développement du numérique qui couvre désormais toutes les facettes de l'information et de la communication. La « convergence numérique » a en effet permis, en quelques années, à tous les types de contenus informationnels (textes, vidéos, images, musiques, etc.) d'être produits, lus et transmis par des appareils et des réseaux identiques alors qu'ils exigeaient auparavant des outils et canaux spécialisés. De nos jours, (presque) tous les textes produits passent à un moment de leur existence par le support numérique, ce qui facilite l'interaction avec les outils du TAL, que ce soit au moment de l'écriture (traitement de texte, aide à la rédaction, correction orthographique, reconnaissance du texte dicté), de la traduction (traduction automatique ou assistée par ordinateur), de la transmission ou de la diffusion (via les réseaux), de l'archivage (indexation des bases de données) ou de l'accès ultérieur à cet archivage par l'intermédiaire de moteurs de recherche.

Il y a quelques années, ces « technologies du langage » qui analysent ou génèrent automatiquement des énoncés langagiers étaient inconnues du grand public. Aujourd'hui, elles font partie de la vie de tous les jours et se sont banalisées au

point qu'on les utilise naturellement, parfois même sans en avoir conscience : on les retrouve dans la voix du GPS, dans l'assistant personnel du téléphone intelligent que l'on interpelle oralement (« dis Siri » ou « OK Google »), dans les outils de traduction automatique, dans les systèmes de dialogue qui répondent à vos questions sur les sites de commerce en ligne, dans l'outil de reconnaissance vocale auquel on dicte un texte, etc. Ces différents services sont le résultat de l'application d'algorithmes, de modèles statistiques et de ressources linguistiques formalisées à différents types de données linguistiques de la vie quotidienne.

Percevant l'intérêt potentiel de ces outils pour la promotion d'actions de politique linguistique, le Service de la langue française de la Communauté Wallonie-Bruxelles a soutenu diverses initiatives allant dans ce sens. Le Centre de traitement automatique du langage (Cental) de l'UCL a ainsi été sollicité pour mettre au point des applications de TAL originales. Nous présenterons dans cet article deux projets réalisés dans ce cadre. La Section 1 détaillera d'abord le projet Recto/Verso, un logiciel capable d'adapter automatiquement un texte depuis l'ancienne orthographe vers la nouvelle. Il s'agit d'un exemple typique de l'usage des technologies de la langue pour promouvoir une politique linguistique particulière, à savoir la réforme de l'orthographe. La Section 2 montrera quant à elle comment les technologies de la langue peuvent aussi se révéler utiles pour promouvoir des actions concrètes, et en l'occurrence rendre les documents administratifs plus accessibles pour le citoyen. Le projet AMesure vise en effet à développer une plateforme d'évaluation des textes administratifs, capable de mesurer la difficulté de ce type de textes et d'identifier les problèmes de rédaction (phrases complexes, vocabulaire technique, etc.) qu'ils peuvent comporter. Le but de ce projet est d'aider les rédacteurs de textes administratifs à rendre leurs productions plus lisibles. Nous terminerons cette contribution en proposant de nouvelles pistes pour une intégration des outils de TAL au sein des politiques linguistiques francophones.

1. Recto/Verso

Le moins que l'on puisse dire, c'est que la « nouvelle orthographe » a déjà fait couler beaucoup d'encre... Issues d'un rapport du Conseil supérieur de la langue française (Paris) et publiées au Journal officiel de la République française le 6 décembre 1990, les rectifications orthographiques ont reçu l'aval de l'Académie française et des organismes de politique linguistique du Québec, de la Suisse et de la Belgique. Le but de ces rectifications est essentiellement de corriger un certain nombre d'aberrations du code écrit (pourquoi écrire *chariot* avec un *r* et *charrette* avec deux ?) et à systématiser certaines règles (l'usage des traits d'union dans les déterminants numéraux composés, le pluriel des noms composés comportant

un verbe, etc.). Si certains y voient une œuvre de rationalisation qui devrait être profitable à l'enseignement et de manière plus générale à la maîtrise de la langue française (et donc indirectement à son rayonnement), d'autres s'opposent à ce qu'ils qualifient de « nivèlement par le bas », de réforme délétère pour la langue française, risquant de lui faire perdre « son âme ». La particularité de cette réforme est qu'elle a été proposée... sans être imposée, contrairement à ce qui s'est passé pour d'autres langues. L'usage devrait choisir disait-on. Force est de constater que presque 20 ans plus tard, l'usage hésite toujours et que son adoption reste partielle, même si elle est officiellement l'orthographe « prioritairement enseignée » en Fédération Wallonie-Bruxelles (depuis 2008) et qu'elle a fait son apparition dans les dictionnaires (plus souvent citée comme variante orthographique qu'utilisée dans les entrées du dictionnaire), dans certains manuels scolaires (suscitant en France de grands émois) et qu'elle est aujourd'hui acceptée par la plupart des logiciels de correction orthographique (Word, Cordial, Antidote, Prolexis, etc.).

À ce propos, l'arrivée de la nouvelle orthographe dans les correcteurs orthographiques, dont celui de Word, a été un évènement très important pour l'adoption de celle-ci par un large public. En effet, comment la nouvelle orthographe pourrait-elle s'imposer dans l'usage, si les outils d'aide à la rédaction - qui eux sont omniprésents - ne la reconnaissent pas ? C'est un indice qui rappelle à quel point les technologies sont aujourd'hui un vecteur de diffusion de la langue et par la même occasion, un vecteur de diffusion potentiel pour certaines politiques linguistiques. Elles représentent certes une opportunité, mais également un point de passage obligé : l'évolution de la langue ou de son code écrit passe aussi par l'adaptation des technologies qui nous servent tous les jours à la manipuler. C'est un fait qui est devenu presque aussi important que l'adaptation des manuels scolaires et des curriculums d'enseignement...

C'est dans ce contexte que le projet Recto/Verso a vu le jour en 2008, à l'occasion de la Semaine de la langue française en fête. Cette action a été réalisée avec de grands acteurs de la presse nationale francophone belge (*Le Soir*, *La Libre Belgique*, *Vers l'avenir*, *La Dernière Heure*) afin de démontrer que les rectifications orthographiques de 1990, loin de remettre complètement le système en cause, proposent des rectifications ciblées, qui ne concernent qu'un nombre minoritaire de formes. Aussi surprenant que cela puisse paraître, dans un article de presse de taille moyenne, il n'y a en général qu'une ou deux modifications (quand il y en a).

Le principe était simple : un bouton marqué « recto/verso » pour « rectification orthographique / version originale » avait été ajouté sur les éditions en ligne de ces quotidiens (Beaufort *et al.*, 2009). Il permettait d'obtenir en un clic une version transcrite en nouvelle orthographe de chaque article. Pour réaliser cette opération, le texte était envoyé sur un serveur qui le convertissait et le renvoyait au journal

pour affichage. En plus de la transcription, une glose explicative était associée à chaque modification pour expliquer la règle correspondant à chaque modification. Ce bouton est resté plusieurs mois sur les éditions des différents journaux et au total, des millions de « transcriptions » automatiques ont été réalisées.

Le logiciel en question ne devait pas « simplement » remplacer une forme (de l'ancienne orthographe) par une nouvelle. L'application des rectifications nécessite en effet de prendre en compte le contexte linguistique (par exemple la suppression de l'accent circonflexe n'a pas lieu dans certains contextes, l'accord des noms composés nécessite de vérifier le nombre, etc.). Le logiciel recto/verso utilisait donc en plus d'un dictionnaire de rectifications, un étiqueteur grammatical et un programme de levée d'ambiguïtés pour assurer l'exactitude des rectifications.

Cette opération a offert une tribune particulièrement visible aux défenseurs de la nouvelle orthographe, puisque du jour au lendemain une partie importante de la presse francophone belge était passée (virtuellement) à la nouvelle orthographe. Cela a permis de :

- rationaliser le débat : en montrant que les changements sont ciblés, qu'il ne s'agit pas d'une remise en cause complète de la langue, contrairement à ce que les détracteurs voudraient laisser penser ;
- diffuser un outil pédagogique permettant de découvrir la nouvelle orthographe en contexte, avec des explications simples associées à chaque modification.

Le service de transcription automatique Recto-Verso est toujours disponible en ligne. Il permet à tout utilisateur de copier/coller son texte et d'obtenir une version adaptée automatiquement. Le résultat de la conversion présente les formes rectifiées en couleurs et soulignées. Quand on passe la souris sur l'une de ces formes, on obtient en superposition sur le texte, une zone de commentaire qui détaille la règle utilisée pour produire la forme rectifiée. Pour des raisons pédagogiques, le système signale aussi (mais dans une autre couleur) les formes qui constituent des règles d'exceptions. Par exemple, les accents circonflexes sur *i* et *u* tombent dans la nouvelle orthographe, mais se maintiennent pour les terminaisons verbales du passé simple, du subjonctif et dans un petit nombre de cas d'homonymie (*dû*, *sûr*, *mûr*, *jeûne*, etc.).

2. AMesure

2.1. Les initiatives pour la simplification du langage administratif

Le projet AMesure¹ a, de son côté, été développé pour rencontrer les préoccupations de diverses administrations francophones quant à l'accessibilité des documents qu'elles produisent. Cette problématique n'est pas neuve et s'inspire des efforts

de la sphère anglophone pour promouvoir l'utilisation d'un anglais simplifié (*Plain English*) dans les administrations. Le mouvement *Plain English* s'est développé aux États-Unis dans les années 70, d'abord grâce à Richard Nixon qui décréta que le registre fédéral américain devrait désormais être lisible pour l'homme de la rue, puis sous l'action de Ronald Reagan qui, en 1978, légiféra afin de rendre les règlements gouvernementaux plus simples à comprendre. En Angleterre, la *Plain English Campaign* a été lancée depuis 1979 dans le même but : aider les ministères et autres institutions à émettre des publications accessibles au plus grand nombre. Cette entreprise privée, fondée par Chrissie Maher, a développé le label *Crystal Mark* qui certifie qu'un document est écrit en langage clair. De façon générale, ces diverses initiatives concernant la lisibilité des documents officiels semblent avoir produit des résultats tangibles. Ainsi, alors que le *American Council of Life Insurance* rapportait que, en 1975, 43% des acheteurs d'une assurance vie exprimaient des difficultés à lire leur police, ce pourcentage était descendu à 33% en 1982. Cet exemple confirme l'intérêt de simplifier les textes administratifs et juridiques à destination du grand public.

La sphère francophone n'est pas moins préoccupée de ces questions d'accès aux documents administratifs et juridiques. Au Québec, tout d'abord, le Secrétariat national à l'alphabétisation publie dès 1994 un guide pour la formation à un style clair et simple. L'année 2000 marque l'adoption de la Loi sur l'administration publique qui vise à améliorer le service aux citoyens. Cette décision politique a donné l'impulsion à une série de recherches menées par le groupe Rédiger de l'Université de Laval (Groupe Rédiger 2002). Ce groupe a notamment conduit une étude visant à identifier les obstacles à une communication claire (cf. Section 2.2) et a participé à la publication, dans le cadre d'une collaboration franco-québécoise, de la brochure *Rédiger simplement - Principes et recommandations pour une langue administrative de qualité* (par le Gouvernement du Québec en 2006). La Suisse romande s'est également préoccupée de ces questions, notamment au moyen de la Commission de rédaction.

En France, la question de la simplicité du langage administratif a été rendue plus prégnante avec la modernisation et le développement de l'informatique et d'Internet. En effet, la relation entre administrations et administrés est de moins en moins directe et s'ancre dès lors davantage sur des documents écrits. Reste à s'assurer que ces documents soient accessibles à tous, dans la lignée de l'exigence républicaine : le principe d'égalité. Pour ce faire, plusieurs organismes ont été créés successivement en France. La Commission pour la simplification des formalités (Cosiform) est instituée en 1990 et lance, en 1996, un groupe de travail sur la lisibilité des formulaires. Ses travaux seront poursuivis au sein de la Commission

pour les simplifications administratives (COSA, décret du 2 décembre 1998), puis du Comité d'orientation pour la simplification du langage administratif (Cosla) en 2001. Les actions du Cosla, en particulier, se sont portées sur les questions de simplification du langage administratif. Ce comité est intervenu pour réécrire les formulaires administratifs², a proposé un *Guide pratique de la rédaction administrative*, un lexique des termes administratifs (Cosla, 2004), ainsi qu'un logiciel d'aide à la rédaction administrative. Ce logiciel, appelé LARA, repère les formulations obscures du langage administratif et « propose des solutions de rechange en langage courant, simple et actuel. Même chose si le terme ou l'expression qu'un rédacteur utilise est trop relevé, trop technique, trop juridique. Si la phrase est trop longue (40 mots), il le signale également. Le rédacteur, ensuite, a le choix de faire ce qu'il veut. » (Langlois, 2002 : 5). Le Cosla a toutefois été dissout et remplacé par la Direction générale de la modernisation de l'État (2005), puis par le Secrétariat général pour la modernisation de l'action publique (SGMAP, 2012), dont les missions étaient davantage dédiées à la simplification des procédures administratives que celles du langage administratif. En 2017, le SGMAP a laissé la place à deux directions interministérielles, la DITP et la DINSIIC.

Enfin, en Belgique, la simplification administrative a été défendue à trois niveaux de pouvoir différents. Au niveau fédéral, c'est l'Agence pour la simplification administrative (ASA) qui vise, depuis 1998, à rendre les procédures administratives plus efficaces. La simplification du langage administratif ne constitue toutefois pas une priorité pour l'ASA. En Wallonie, le Commissariat EASI-WAL a été fondé en 2004 par la Fédération Wallonie-Bruxelles et il a été investi de 4 missions principales, dont l'une est de veiller à l'amélioration de la lisibilité des textes administratifs. En 2013, il fusionne avec le Centre d'Expertise du Numérique (Cepige) pour devenir eWBS (*e-Wallonie Bruxelles Simplification*) et ses missions se réorientent principalement vers la simplification et la personnalisation des procédures administratives, notamment grâce au numérique. Enfin, très actif sur les questions de simplicité du langage administratif, le Service de la langue française de la Fédération Wallonie-Bruxelles est à l'origine, depuis plus de 25 ans, de diverses mesures en faveur d'une communication claire et efficace avec le citoyen. Dès 1993, Christiane Strauven publie dans la revue *Français et Société* un numéro spécial sur le langage administratif, qui est soutenu par la FWB (Strauven, 1993). Une brochure intitulée *Ecrire pour être lu : Comment rédiger des textes administratifs faciles à comprendre ?* qui regroupe des conseils de rédaction claire, a également été éditée à plusieurs reprises par la FWB, la dernière version datant de 2017. Plus récemment, une Commission *Lisibilité* a été mise en place au sein du Conseil de la langue française et de la politique linguistique (CLFPL), laquelle a soutenu le développement du projet AMESURE décrit ci-dessous.

2.2. Le projet AMesure : une plateforme de lisibilité en ligne

Le Groupe Rédiger est notamment l'auteur d'une étude sur la lisibilité de différents genres d'écrits destinés au grand public (lettres, formulaires, courriels, formulaires électroniques, sites, etc.). Cette analyse de corpus a été complétée par des enquêtes de terrain visant à évaluer la compréhension et l'utilisabilité de ces documents, ainsi qu'à établir les contraintes de rédaction de ces documents. Il en est notamment ressorti que :

1. Les communications écrites sont peu ou mal adaptées aux citoyens ;
2. Les chaînes de production des textes administratifs ne facilitent pas l'implantation de la simplification (Clerc, 2009 : 32).

Sans entrer ici dans les détails de ces entraves à l'implantation des procédures de simplification, lesquelles sont pourtant bien documentées dans les guides de rédaction simple mentionnés plus haut, il semble effectivement que les documents administratifs continuent de poser des problèmes de compréhension. Le développement du projet AMesure est parti de ce constat et de l'espoir que l'automatisation des conseils disponibles dans les différents guides de simplification simple puisse aider les rédacteurs de textes administratifs à mieux mettre en œuvre ces conseils dans leurs productions. Le principal résultat du projet AMesure consiste en une plateforme web permettant d'évaluer la difficulté des textes administratifs à l'aide de plusieurs outils issus du TAL et du domaine de la lisibilité.

Si l'on ne reviendra pas sur le domaine du TAL, il convient de présenter brièvement le domaine de la lisibilité avant de se focaliser sur le projet AMesure. La lisibilité est un domaine qui est né aux États-Unis dans les années 1920 et a rencontré un vif succès dans la culture anglo-saxonne, en particulier après qu'ont été publiées les formules de Flesch et de Dale and Chall (Conquet, Richaudeau, 1973). Celles-ci reposent sur des approches simples : la difficulté d'un texte est assimilée à la longueur des mots et à la longueur des phrases. Ces formules ont pourtant été largement utilisées, que ce soit pour étalonner les matériaux de manuels scolaires, pour réduire la difficulté des articles dans les quotidiens de presse ou pour rendre les textes administratifs et techniques plus simples. Dans le domaine francophone, il faut attendre 1956 pour qu'André Conquet mette en exergue le domaine de la lisibilité dans son ouvrage du même nom (Conquet 1971). Peu après, la première formule de lisibilité pour le français est proposée par Kandel et Moles (1958), même s'il s'agit simplement d'une adaptation de la formule de Flesch. Henry (1975) propose la première formule de lisibilité spécifique au français, qui est conçue pour le milieu scolaire. Récemment, le domaine a connu un renouveau sous l'effet des techniques issues du TAL et de l'intelligence artificielle, lesquelles permettent

d'analyser automatiquement un plus grand nombre de dimensions textuelles et des caractéristiques textuelles plus complexes³ (ex. fréquence des mots, polysémie, structures syntaxiques, expressions polylexicales, etc.). François et Fairon (2012) ont ainsi développé la première formule de lisibilité computationnelle pour le français langue étrangère ou seconde, qui a servi de base au projet AMesure. Les approches actuelles en lisibilité permettent en effet de détecter, dans les textes, différents phénomènes linguistiques susceptibles d'engendrer des difficultés de lecture et de compréhension.

Le projet AMesure a connu plusieurs phases successives. Tout d'abord, nous avons mis au point une formule de lisibilité spécialisée pour les textes administratifs, étant donné que ni celle de Henry (1975), ni celle de François et Fairon (2012) n'avaient été pensées pour des textes de spécialité. Elle propose à l'utilisateur une indication globale de la difficulté du texte, définie sur une échelle de 1 (très simple) à 5 (très difficile). Dans une seconde phase du projet, nous avons poursuivi nos recherches en développant une plateforme, appelée AMesure, librement accessible en ligne⁴. La plateforme offrait, dans un premier temps, uniquement une indication globale de la difficulté du texte, définie sur une échelle de 1 (très simple) à 5 (très difficile) ainsi qu'une série d'indicateurs précisant la difficulté de différentes dimensions d'un document administratif, à savoir la proportion de mots absents du Français Fondamental (Gougenheim *et al.*, 1964), le nombre moyen de mots par phrase, la proportion de conjonctions/pronoms, un indice du degré de personnalisation du texte et une mesure de la cohérence moyenne du texte. Plus de détails concernant les aspects scientifiques de ces deux premières étapes ont été publiés par ailleurs (François *et al.*, 2014).

2.3. Le projet AMesure : vers la simplification automatique des textes

Lors de la troisième étape du projet, le logiciel a connu une refonte assez importante. En effet, suite à des premiers retours de la part de rédacteurs de textes administratifs, il est apparu que se limiter à des indicateurs numériques ne suffisait pas pour aider les rédacteurs à améliorer la clarté de leurs productions. Nous avons donc adopté une autre approche, à savoir détecter directement les phénomènes difficiles à lire dans les textes administratifs. Ces phénomènes peuvent correspondre à des mots rares, des abréviations obscures ou trop denses, des structures de phrase plus complexes, etc. Cette nouvelle approche s'inspire désormais moins des travaux en lisibilité que du domaine de la simplification automatique de textes (SAT). L'objectif de la SAT consiste à modifier automatiquement un texte ou un énoncé afin de le rendre plus accessible à la lecture, tout en conservant son sens aussi

intact que possible. On peut structurer le domaine de la SAT à l'aide de deux grands axes : (1) simplifications syntaxiques vs simplifications lexicales et (2) les approches par règles vs les approches statistiques. La simplification syntaxique consistant essentiellement à simplifier les structures syntaxiques, elle a été soit abordée à l'aide de règles linguistiques appliquées au résultat d'un analyseur syntaxique automatique, soit via des techniques d'apprentissage automatique (modèles statistiques) appliquées à des corpus parallèles (version originale et simplifiée). La simplification lexicale, quant à elle, consiste généralement à remplacer les mots détectés comme complexes par des synonymes plus simples et adaptés au contexte d'origine. Pour la détection des mots complexes, les approches les plus simples sont fondées sur la fréquence, tandis que d'autres combinent plusieurs variables lexicales au sein de modèles statistiques. La génération de synonymes peut se faire sur la base de ressources linguistiques, telles qu'un thésaurus ou une ressource de synonymes collaborative⁵. D'autres chercheurs préfèrent obtenir des candidats synonymes automatiquement à partir de textes, en utilisant des corpus parallèles ou en utilisant les modèles statistiques.

Le projet AMesure vise à s'inspirer de ces techniques pour indiquer automatiquement des pistes pour la simplification manuelle de textes administratifs. Ces pistes sont issues à la fois des conseils des guides de rédaction simple publiés par les diverses administrations francophones, mais également des connaissances en lisibilité et des études en psycholinguistique décrivant les facteurs influençant le processus de lecture. Actuellement, la dernière version de la plateforme, accessible gratuitement en ligne, détecte trois grands types de phénomènes : (1) les structures syntaxiques complexes (subordonnées et passives), (2) les abréviations et (3) les mots complexes. Pour plus de détails sur les processus de détection de ces phénomènes et leur évaluation, le lecteur peut se reporter à François *et al.* (2018).

Enfin, lors de la quatrième et dernière étape de développement (à ce jour), la plateforme a été enrichie par un ensemble de conseils de rédaction simple, localisés directement au niveau des phrases problématiques. Autrement dit, le système ne se contente plus de détecter des phénomènes linguistiques dans le texte, mais il est également capable d'analyser, au niveau de la phrase, l'ensemble des problèmes de lisibilité détectés et de produire un diagnostic adapté en précisant les éléments qui devraient être améliorés. En effet, le système précédent détectait systématiquement les propositions subordonnées et pouvait ainsi mettre en évidence de courtes propositions relatives ou infinitives, qui ne posent en réalité pas de problèmes de compréhension. Après avoir consulté les guides de rédaction simple présentés ci-dessus, nous avons défini un ensemble de règles linguistiques capables de détecter des situations correspondant à des cas problématiques mis en exergue

dans ces guides. Par exemple, les phrases longues comportant plus de 3 subordonnées ou au moins trois structures enchâssées l'une dans l'autre sont détectées par le système.

Au niveau de l'interface, ces modifications se traduisent par l'ajout de bulles infos (sur le côté droit de l'interface) intégrant ces conseils. L'onglet *Analyse des phrases* surligne le texte avec diverses nuances de jaune : plus une phrase est surlignée dans un jaune vif, plus elle comprend de difficultés et plus elle gagnerait à être simplifiée. Cette partie de l'interface offre donc une vue globale du texte, qui devrait permettre d'attirer l'attention des rédacteurs de textes administratifs sur les aspects les plus problématiques de leurs textes. Les autres onglets (accessibles via le menu gauche) permettent ensuite de s'attacher à chacun des phénomènes isolément, par exemple, les mots complexes ou les abréviations.

Pour conclure cette section sur l'outil AMesure, nous avons montré comment un outil de TAL, accessible via Internet, peut non seulement servir à soutenir les politiques de simplification des documents administratifs à destination des citoyens, mais aussi à favoriser leur implantation concrète au sein des ministères. Il est prévu de poursuivre ce projet en analysant plus précisément les pratiques de simplification actuelles dans les administrations belges, à l'instar de ce qui a été fait par le groupe Rédiger, mais également d'évaluer l'impact réel d'AMesure auprès des rédacteurs et de l'améliorer en conséquence.

Conclusion : TAL et politique linguistique

Bien que les deux exemples présentés ci-dessus soient très différents l'un de l'autre, ils ont en commun d'être des outils conçus pour être au service du citoyen et reposant sur les technologies de l'ingénierie linguistique. Accessibles sur Internet, ils permettent de faire des démonstrations linguistiques à très large échelle et ont également une visée pédagogique puisque l'on découvre certains aspects du fonctionnement de la langue au travers de leur utilisation. Les technologies du langage offrent donc de nombreuses opportunités pour sensibiliser le public (grâce à des outils pédagogiques ou à des démonstrations à large échelle) à certains aspects de la langue française qui peuvent par exemple correspondre à certaines préoccupations de politique linguistique (nouvelle orthographe, féminisation, lisibilité, etc.). C'est ce que souligne le Conseil de la langue française et de la politique linguistique de la FWB dans son avis du 8 octobre 2009⁶.

Au-delà des opportunités que représentent les outils de traitement automatique du langage, il faut de manière plus générale prendre conscience de l'importance croissante de ces technologies dans notre vie quotidienne (nous avons par exemple

mentionné la voix du GPS, le correcteur orthographique, la traduction automatique, les systèmes de dialogue, etc.). La bonne représentation du français au sein de ces outils est donc de facto un enjeu important pour le statut de la langue dans la société de l'information. Les langues « compatibles » avec ces technologies (c'est-à-dire les langues pour lesquelles les technologies ont été adaptées) seront nécessairement favorisées par rapport à celles qui sont plus difficilement représentées dans l'univers numérique au sein duquel nous évoluons. Les industries de la langue de nos régions ont un rôle important à jouer et il faut stimuler leurs capacités de recherche et de développement. Il y a en effet tout lieu de souhaiter que la création des ressources linguistiques et des outils informatiques nécessaires au développement des technologies du langage pour le français soit prise en charge par les pays francophones eux-mêmes. Dans cette perspective, stimuler l'ingénierie linguistique dans tous les pays de la francophonie doit être à la fois un objectif économique et un objectif de politique linguistique visant à promouvoir et défendre l'usage de la langue française dans la société contemporaine. Pour conclure en revenant au contexte belge, nous appelons à une concertation des institutions qui promeuvent le développement économique et la recherche industrielle (Région wallonne) avec celles qui soutiennent la recherche fondamentale (FNRS) ou travaillent à la promotion de la langue française (Fédération Wallonie-Bruxelles). C'est un enjeu stratégique pour le développement de nos régions et l'avenir de notre langue.

Bibliographie

- Beaufort, R., Dister, A. Naets, H., Macé, K., Fairon, C. 2009, « Recto/Verso. Un système de conversion automatique ancienne/nouvelle orthographe à visée linguistique et didactique ». In : *Actes de TALN 2009*, Senlis.
- Clerc, I. 2009. « Langue française et écriture citoyenne : l'engagement du Québec en matière de simplification des communications écrites ». In : *La communication avec le citoyen : efficace et accessible ? Actes du colloque de Liège 2009*. De Boeck-Duculot, p. 27-40.
- Conquet, A., Richaudeau, F. 1973. Cinq méthodes de mesure de la lisibilité. *Communication et langages*, 17(1), 5-16.
- COSLA. 2004. *Lexique administratif*. Paris : Le Robert.
- François, T. 2011. *Les apports du traitement automatique du langage à la lisibilité du français langue étrangère*. PhD thesis, Université Catholique de Louvain.
- François, T., Brouwers, L. et Naets, H., Fairon, C. 2014. « AMesure : une formule de lisibilité pour les textes administratifs ». In : *Actes de TALN 2014*, Marseille, p. 467-472.
- François, T., Fairon, C. 2012. « An 'Al readability' formula for French as a foreign language ». In : *Proceedings of the 2012 Conference on EMNLP 2012*, 466-477.
- François, T., Müller, A., Degryse, B., Fairon, C. 2018. « AMesure : une plateforme web d'assistance à la rédaction simple de textes administratifs ». In : *Repères-Dorif*.
- Gougenheim, G., Michéa, R., Rivenc, P., Sauvageot, A. 1964. *L'élaboration du français fondamental (1er degré)*. Didier, Paris.

Groupe Rédiger (dir. I. Clerc). 2002. *Simplification de la langue administrative : les formulaires imprimés et électroniques, les fiches du Répertoire des programmes et services*, Québec : ministère des Relations avec les citoyens et de l'Immigration.

Henry, G. 1975. *Comment mesurer la lisibilité*. Labor, Bruxelles.

Kandel, L. et Moles, A. 1958. « Application de l'indice de Flesch à la langue française ». In : *Cahiers Études de Radio-Télévision*, 19, p. 253-274.

Langlois, C. 2002. « Le COSLA, une occasion rêvée de simplifier le langage administratif ». In : *L'Actualité terminologique*, vol.35(2), p. 5.

Rehm, G., Uszkoreit, H. (éd.). *La langue française à l'ère du numérique*. Collection de Livres Blancs.

Réseau Meta-Net. Springer. s.d. <http://www.meta-net.eu/whitepapers/e-book/french.pdf> [consulté le 20/01/2018].

Richard B., Anne D., Hubert N., Kévin M., Cédric F. 2009. « Recto/Verso. Un système de conversion automatique ancienne/nouvelle orthographe à visée linguistique et didactique ». In : *Actes de TALN 2009*.

Strauven, C. 1993. *La lisibilité des textes administratifs*. Communauté française Wallonie Bruxelles, Direction générale de la culture et de la communication, Service de la langue française.

Notes

1. <https://uclouvain.be/recto-verso/> [Consulté le 20 avril 2018].

C'est un plaisir de remercier l'équipe du CENTAL qui travaille au développement du projet AMesure et notamment Adeline Müller, Gaëtan Ansotte, Brayan Delmée, Baptiste Degryse, ainsi que le Service de la langue française de la Communauté Wallonie-Bruxelles pour son soutien.

2. Pour un exemple, consulter le site

https://www.fonction-publique.gouv.fr/archives/home20051107/communications/dossiers-presse/archives/cosla_05mars02.pdf. [Consulté le 20 avril 2018].

3. Pour une introduction plus détaillée à ce domaine, le lecteur se reportera à la thèse de François (2011) pour une synthèse détaillée, incluant les travaux en français.

4. <http://cental.uclouvain.be/amesure/> [Consulté le 20 avril 2018].

5. Il s'agit de JeuxdeMots (Lafourcade, 2007), une ressource de synonymes constituée de façon collaborative à l'aide d'un jeu.

6. <https://goo.gl/txwd23> [Consulté le 20 avril 2018].