

INTRODUCTION

Thierry PONCHON
Université de Reims Champagne-Ardenne
et STIH - Paris-Sorbonne (EA 4509)

&

Isabelle LABORDE-MILAA
Université Paris-Est Créteil, Ceditec (EA 3119)

Cet ouvrage constitue les actes du colloque « Sciences du langage et nouvelles technologies » organisé par l'Association des Sciences du Langage le 5 décembre 2009 à la Sorbonne, avec le soutien de l'université Paris-Sorbonne (Paris IV) et de l'équipe *Sens - Texte - Informatique - Histoire* (EA 4509, Paris IV).

Les liens entre la linguistique et les nouvelles technologies sont déjà relativement anciens. C'est ainsi que des réflexions épistémologiques et philosophiques sur les rapports entre information et langage ont présidé à la naissance de l'informatique. Toutes deux ont en commun de travailler sur des données, sinon totalement immatérielles, du moins fortement dématérialisées, au point qu'on a pu en parler comme de sciences de l'esprit¹, comme l'illustre l'entretien de Jean-Guy Meunier à la Maison des Sciences de l'Homme de Paris².

D'autres auteurs, comme Antoine Guillaume³, ont montré l'intérêt des nouvelles technologies pour les sciences du langage, mais aussi à quel point il est illusoire de penser qu'elles puissent se substituer à la réflexion linguistique.

De nombreux chemins ont été balisés depuis, de sorte que les avancées dans les deux domaines – et avant tout dans leur *conjointure* –, méritent d'être portés régulièrement à la connaissance de la

1. http://www.canal-u.tv/producteurs/les_amphis_de_france_5/dossier_programmes/philosophie/sciences_et_philosophie/les_sciences_de_l_esprit

2. www.archivesaudiovisuelles.fr/2/introduction.asp.

3. www.languescultures.fr/fr/_encyclopubbykeyword.asp?motcle=linguistique+informatique

communauté scientifique ; c'était l'objectif du colloque dont ces actes veulent rendre compte : donner un panorama des usages que font les linguistes des nouvelles technologies. Pour autant, ces actes, aussi riches soient-ils, n'en donnent *de facto* qu'une vue parcellaire. Les interactions entre les sciences du langage et les nouvelles technologies sont trop diverses pour qu'on puisse en dresser un état des lieux en un seul volume. Ces pages témoignent bien davantage de la volonté de présenter des études récentes ou en cours pour attester combien les nouvelles technologies peuvent apporter aux sciences du langage, comment ces dernières, en retour, poussent les outils à s'améliorer et, en définitive, combien fructueuses sont les recherches collaboratives et les applications qui peuvent en découler.

Les articles qui composent ce recueil montrent cette synergie entre la recherche linguistique et les nouvelles technologies, aide précieuse pour l'exploitation de ressources, non seulement par la numérisation de corpus oraux et écrits, de dictionnaires, de lexiques, mais aussi par l'élaboration d'outils de traitement et d'analyse. De plus, elles facilitent la communication, la coopération, le partenariat et les échanges scientifiques (forums, visioconférences...).

L'exemple spontané que l'on peut donner pour symboliser à la fois l'enrichissement que procurent les nouvelles technologies et la progression dans leur compétence et leur efficacité au regard de la langue est dans l'aide à la rédaction. La première génération de correcteurs orthographiques non intégrés vérifiait simplement l'existence d'un « mot » dans une liste lexicale préétablie. La deuxième génération ne prenait en compte que l'environnement immédiat. En dépassant le stade de l'analyse (syntaxique) locale, les outils actuels de « correction orthographique » ont beaucoup progressé et continuent d'évoluer. Ils s'intègrent de mieux en mieux aux logiciels de traitement de texte ou d'analyse de données, voire de messagerie. À l'analyseur orthographique de départ, complété par la suite par un analyseur flexionnel en contexte rapproché, se sont ajoutés de multiples dictionnaires et les analyses syntaxiques (bien perfectibles encore) s'appliquent à un contexte de plus en plus large. Les correcteurs orthographiques ont ainsi évolué vers l'analyse phrastique. Devenus indispensables, les correcteurs grammaticaux « intelligents » se développent désormais à partir d'une conception *fractale* de la langue, d'un monde des possibles : tout en traitant l'information avec une remarquable célérité, ils s'appuient sur des critères définis par le scripteur lui-même et

suggèrent plus qu'ils n'imposent. La fonction pédagogique contre-balance l'accroissement exponentiel des possibilités d'erreurs.

Cette application permet de mesurer que depuis les premières approches – il y a plus de cinq décennies –, les progrès permis par les nouvelles technologies dans l'analyse du langage naturel sont considérables. Grâce notamment au traitement automatique des langues naturelles et à ses applications, il est permis désormais de s'appuyer sur une modélisation et sur des analyses fines pour décrire certaines parties de la linguistique.

Dans les domaines de la phonétique et de la phonologie, les « nouvelles technologies linguistiques » ont abouti à la transcription automatique de la parole (dictée automatique, informatique ou numérique par reconnaissance vocale), à la réalisation de systèmes de segmentation et de synthèse de la parole, à la création et à l'exploitation de systèmes de codage, comme Philippe Caron en décrit un exemple dans son article sur un projet (mené avec Michel Morel) d'aide à la diction des vers français.

Tirant parti des travaux sur les lemmatiseurs, les analyseurs, les générateurs de mots, les recherches axées sur les aspects morphologiques des langues bénéficient aussi du soutien des nouvelles technologies. L'article de Sonia Branca-Rosoff et Serge Fleury sur l'examen des emplois respectifs de deux tiroirs verbaux (futur simple et futur périphrastique) à partir d'un imposant corpus oral, l'atteste bien. L'appui fourni par un logiciel de textométrie intégrant un système d'étiquetage automatique des catégories grammaticales des mots avec lemmatisation permet, entre autres, de prouver que le sentiment commun d'une emprise du futur périphrastique à l'oral est erroné, tout particulièrement dans les emplois modalisants. En matière de syntaxe, des analyseurs, des générateurs de phrases et des extracteurs automatiques sont opérationnels. C'est dire que l'analyse syntaxique a bien tiré profit des nouvelles technologies et ne cesse de le faire.

Si les travaux en dictionnaire ne peuvent plus se concevoir sans user du traitement automatique des langues, nombreux aussi sont ceux qui y ont recours en lexicologie. L'article d'Emmanuel Cartier et Jean-François Sablayrolles en est un parfait exemple, article dans lequel il est confirmé que les apports de la technologie sont nonpareils et essentiels pour repérer et jauger efficacement l'émergence et l'expansion des néologismes. En revanche, parce que le sens est particulièrement complexe à décrire et à formaliser, les réalisations dans le domaine sémantique demeurent moins nombreuses et ne concernent que des applications limitées où l'analyse se réduit à un domaine restreint. Bien qu'ardue, l'élabo-

ration d'un analyseur sémantique intégrant la totalité de la langue s'avère être un enjeu crucial pour les recherches à venir.

Le programme d'Analyse Automatique du Discours (imaginé en 1966 et publié en 1969 par Michel Pêcheux) a constitué une des premières tentatives notables pour construire des procédures capables de restituer la trace de la structure invariante des discours avec le secours de l'informatique. Le dispositif mis en place était composé d'une phase manuelle de délimitation de tous les énoncés élémentaires des corpus et d'une phase de traitement informatique, comportant un algorithme de comparaison de contenus lexicaux en lien avec des indicateurs morphosyntaxiques, et un algorithme construisant des chaînes relationnelles de proximité permettant de regrouper des séquences discursives. Les données récupérées et traitées aboutissaient à des listes d'énoncés et des listes de relations entre ces derniers. Toutefois, la spécificité de ce programme fut de prendre appui sur les théories harrissiennes et d'exclure délibérément – entre autres – la narratologie et le langage ordinaire. Il n'en reste pas moins que cette recherche a – *mutatis mutandis* – débouché sur l'analyse du discours assistée par ordinateur.

Depuis, les dispositifs informatiques ont eu pour effet d'ouvrir les champs de recherche, d'élargir l'étude des processus discursifs et d'appréhender le corpus comme un lieu de questionnement sur la construction de problématiques discursives. Qui plus est, certains programmes et outils informatiques offrent désormais l'opportunité aux linguistes de se focaliser sur des aspects dans l'analyse du discours qu'il eût été impossible d'envisager auparavant. Les supports informatiques servant à identifier les opérations textuelles lors de la composition d'écrits, que décrit Thierry Olive dans sa contribution, lui permettent de s'appesantir sur le rôle – naguère négligé – des pauses, et de pouvoir présenter ainsi non pas tant une analyse de la production de texte qu'une analyse en temps réel. Ce paramètre, qu'il nomme « discours temporel », offre une vision nouvelle et enrichissante quant à la génétique textuelle, et une meilleure compréhension des processus rédactionnels et cognitifs en jeu.

Dans une autre perspective, en s'intéressant à la « communication facilitée », qui consacre l'ordinateur comme un outil d'aide à la production d'écrits par des enfants autistes, Christian Hudelot présente une analyse des aspects syntaxico-sémantiques, lexicaux et phraséologiques du discours s'appuyant sur une conception dialogique de l'interlocution.

Les liens qui se construisent entre nouvelles technologies et sciences du langage s'avèrent donc propices au développement de

connexions avec d'autres champs disciplinaires. La contribution de François-Xavier Bernard et Michael Baker propose une réflexion sur les interrelations et les interactions dans la communication médiée par ordinateur entre des lycéens. Leur article envisage les situations d'apprentissage coopératif instrumenté et les contenus des interactions argumentées ; il approfondit les processus d'intégration des technologies éducatives et leurs pratiques. Ce faisant, les auteurs ne sacralisent pas l'outil et son dispositif spécifique : ils mettent en valeur le fait que, sans apprentissage fin et contextualisé, les technologies peuvent produire des effets décevants ou décalés.

De même, l'étude faite par Ivan Šmilauer sur un dispositif de diagnostic automatique des erreurs dans l'apprentissage du français langue étrangère (illustrée par l'analyse des erreurs de déclinaisons en tchèque à partir d'une application en accès libre) met en avant l'intérêt que peut procurer l'intégration du traitement automatique de la langue (TAL) dans les outils d'enseignement des langues assisté par ordinateur (ELAO). La formalisation induite par le TAL et son adaptation font porter un regard nouveau sur l'acquisition d'une langue étrangère et sur l'ELAO, mais confirment aussi que l'intégration du TAL bâtit un pont entre l'ELAO et les sciences du langage.

Ainsi, les recherches présentées dans ces actes donneront au lecteur une idée de la manière dont émerge des travaux en sciences du langage une acuité scientifique qui confirme au besoin leur importance dans l'économie de la connaissance et leur légitimité sociale et culturelle.