
Notes de lecture

Rubrique préparée par Didier Schwab

Université Grenoble Alpes, Laboratoire d'Informatique de Grenoble

Erika Lombart. L'implicite sur les réseaux sociaux. Entre les lignes des forums de discussion. L'Harmattan. 2022. 271 pages. ISBN 9782140294303.

Lu par **Gaëlle Baudry**

Chargée de marketing et communication de l'innovation à Berger-Levrault.

Erika Lombart offre un regard à la croisée de la rhétorique et de la pragmatique, en plongeant au cœur de l'univers du web 2.0, dit aussi web social ou participatif. Dans ce cyberspace de production, elle s'intéresse au plus près d'utilisateurs-internautes engagés sur des réseaux sociaux, en retraçant une partie de leurs microcontenus rapides, authentiques, et innombrables. De la sorte, elle souhaite questionner le lien supposé entre les formes de l'implicite non conventionnel (NC) et l'intensité émotionnelle prenant part dans leurs échanges. L'auteure débute par une présentation succincte de l'état de l'art, suivie d'une analyse terrain où elle adopte une approche originale du concept d'implicite NC, en l'appréhendant tel un phénomène global évoluant au gré d'interactions conversationnelles entre les internautes participants. Elle s'attelle plus précisément à déconstruire leurs conversations ayant lieu sur le forum Doctissimo, d'une part pour en saisir localement les diverses formes linguistiques constitutives, et d'autre part pour en comprendre globalement leurs utilisations dans la dynamique de construction du réseau.

Le premier chapitre reprend les théories conversationnelles appliquées à la communication écrite médiée par ordinateur. L'implicite et ses multiples formes sont positionnés, ainsi que la place et les rôles des participants nécessaires à la formalisation de l'échange. Pour ce faire, la clarification s'instaure entre forme conventionnelle telle que le présupposé par ancrage direct, et forme non conventionnelle telle que le sous-entendu par ancrage indirecte (antiphrase, ironie...). Erika Lombart pose ensuite les contours de ses outils d'analyse, menant à l'interprétation complexe des énoncés. Pour ce faire, elle s'attache à dresser la différence entre acte illocutoire et effet perlocutoire. Le premier réfère à l'intention du locuteur et nécessite reconnaissance pour exister. Le second en est la conséquence, ou plutôt les possibles conséquences (faire faire/croire/sentir/savoir) en fonction du champ émotionnel contextuellement contraint propre à chaque destinataire à réception du discours. Elle liste également nombre de maximes

conversationnelles, dont l'identification permet d'appréhender au mieux les jeux relationnels entre interlocuteurs. Enfin, elle porte une attention particulière à une pluralité de figures implicites en prenant soin de qualifier leurs constructions et leurs finalités discursives. Sept d'entre elles sont retenues pour l'analyse, car entretenant un lien avec une forme NC : la litote, l'hyperbole, l'allusion, la métaphore, la métonymie, le trope illocutoire et l'ironie.

Dans le deuxième chapitre, l'auteure met en tension ces réflexions discursives avec les caractéristiques inhérentes à leur environnement spécifique de production numérique. Partant du principe que le discours est composite, elle adopte une approche croisée mêlant une analyse linguistique portant sur la production et l'interprétation du sens, avec une approche extralinguistique relative aux spécificités technologiques d'expression.

Ainsi, elle investit le cadre particulier des forums de discussion, où se mêlent une communication à la fois thématique et sociale, interpersonnelle et de masse, quasi synchrone et différée, révélatrice de détails intimes sous couvert de pseudonymat. Par ailleurs, sa viabilité et sa longévité se fondent sur une dynamique communautaire virtuelle, dont les échanges sont modérés a posteriori. Erika Lombart s'intéresse alors au rôle de l'implicite NC dans l'expression des émotions lors de ces e-communications, où l'absence de contexte physique partagé ne permet pas l'usage du non-verbal. Elle met en exergue diverses stratégies usitées pour pallier ce manque : émoticône, typologie, ponctuation, disposition spatiale, acronyme, interjection, didascalie. Pour ce faire, elle se fonde sur un corpus riche et contrasté d'échanges entre membres âgés de 18 à 44 ans majoritairement, issus de forums relatifs à la santé et à la psychologie. Elle s'immerge ainsi entre les lignes des forums portant sur les thématiques de la maladie d'Ehlers Danlos, de la vision abîmée par l'éclipse solaire, de la hernie, des vaccins pédiatriques, du cancer de proches, et de la contraception féminine. Les émotions recueillies oscillent entre colère, tristesse et peur, et peuvent être exprimées envers des membres de la communauté ou des extérieurs, avec plus ou moins d'intensité en fonction des sujets et des temporalités conversationnelles.

En préambule à l'analyse, Erika Lombart rend compte de la difficulté à qualifier l'écart entre la réalité et son expression par les internautes, l'invitant dès lors à ajuster quelque peu son protocole original afin de rester la plus fidèle possible aux contenus conversationnels. Les premiers résultats s'appuient sur une vision globale puis segmentée des fils de discussion des six forums étudiés. Ils confirment sa première hypothèse de recherche, qui lie la croissance de l'intensité émotionnelle avec l'amplification de l'usage des formes de l'implicite NC, toutes formes confondues. Ensuite, l'auteure revient sur les indices ayant favorisé la reconnaissance de l'implicite NC dans ces forums, et la compréhension de leurs constructions. D'une part, elle se penche sur leurs formes et sur leurs fréquences via un travail de catégorisation spécifique reprenant notamment le lexique, la sémantique, la syntaxe, la pragmatique. D'autre part, elle s'intéresse à leur signalisation via une analyse dense de stratégies mises en place par des internautes, par l'usage de divers émoticônes ou acronymes, et par le type de ponctuation et le type des phrases émises. À terme, elle propose une modélisation croisant les

constructions des formes de l'implicite NC en conversation avec leurs fréquences. Elle présente ainsi un travail enrichissant les premiers savoirs théoriques sur le sujet, et propose également une nouvelle catégorisation portant particulièrement sur l'ironie dite par la fausse ingénuité. Afin d'appréhender ce phénomène plus globalement, elle enrichit ces résultats d'une approche combinatoire mettant en relation les formes de l'implicite NC telles des composantes combinables constitutives de cette dynamique. L'auteure poursuit alors en émettant une hypothèse sur le rôle des intentions perlocutoires dans la gestion de la communication. Plus précisément, elle met en évidence l'usage privilégié de certaines formes de l'implicite NC en fonction de la nature des émotions véhiculées dans le fil de discussion, s'exprimant tantôt au service d'une modalité relationnelle, tantôt au service d'une modalité se voulant efficace ou persuasive.

Pour conclure, Erika Lombart présente les limites et les perspectives de son travail. Elle insiste entre autres sur le biais émotionnel induit par le corpus usité penchant sur la dimension exclusivement négative, justifiée par la thématique sanitaire et sociale portée par les forums étudiés. La base de données constituée, relatives aux formes de l'implicite NC, offre tout de même de belles perspectives de réutilisation. En effet, celle-ci pourra enrichir les outils de traitement automatique des langues, dont une des difficultés persistantes vise à cerner les formes de l'implicite NC dans les discours. Par ailleurs, sa méthode originale pourrait inspirer d'autres chercheurs souhaitant approfondir le sujet ou mener des études comparatives auprès d'autres populations, d'autres environnements web ou encore d'autres thèmes fédérateurs.

Helena Moniz, Carla Parra Escartin (Eds). Towards Responsible Machine Translation, Ethical and Legal Considerations in Machine Translation. Springer. 2021. 233 pages. ISBN 978-3-031-14688-6.

Lu par **Eunika Mercier-Laurent**

Chercheur associé URCA/CRéSTIC, Chair IFIP Technical Committee Artificial Intelligence

Ce livre est un recueil de contributions des chercheurs ayant pour objectif de réfléchir sur les différents aspects de la traduction automatique responsable. Il ouvre un dialogue entre les disciplines comme la philosophie et le droit et propose un large éventail de sujets de réflexion. Il couvre les aspects liés au développement des systèmes de traduction automatique, ainsi que leur utilisation dans différents scénarios et l'impact sociétal qu'ils peuvent avoir. Ce texte s'adresse aux étudiants et aux chercheurs en linguistique, en traduction, en traitement du langage naturel, en philosophie et en droit, ainsi qu'aux professionnels travaillant dans ces domaines, comme les traducteurs, les interprètes et les utilisateurs des outils de traduction automatique.

Ce livre est composé de 3 parties. La première considère la traduction automatique du point de vue éthique, philosophique et juridique. La seconde présente trois points de vue sur l'éthique de la traduction automatique perçue par les utilisateurs et la dernière discute quelques éléments de l'impact sociétal de divers traducteurs.

Ce livre est composé de trois parties. La première considère la traduction automatique du point de vue éthique, philosophique et juridique. La seconde présente trois points de vue sur l'éthique de la traduction automatique perçue par les utilisateurs, et la dernière discute quelques éléments de l'impact sociétal de divers traducteurs.

Les auteurs de la première partie soulèvent des points similaires à celles déjà débattues sur l'éthique de l'IA en général, dans les contextes européen, UNESCO-COMEST, World Economic Forum et autres, à savoir la propriété intellectuelle, les droits d'auteur, la protection des données lors de leurs exploitations diverses. Une question intéressante posée dans cette partie concerne la signification des expressions qui peut être perdue lors d'une traduction automatique par des machines qui ne savent pas raisonner. Quelques anciennes questions philosophiques, « est-ce que les machines sont capables de penser ? » ou « qu'est-ce le dialogue artificiel ? » sont encore posées. A. Nousias présente une perspective grecque sur l'essence de la signification et sur la créativité linguistique en général. Cette créativité dépend de plusieurs facteurs, notamment du background de la personne qui s'exprime et de son contexte culturel. Il propose de calculer les corrélations entre les langues et les pays. Toute « artificialisation » mène à la perte d'information. Il est donc important de prendre en considération dans la conception des systèmes de traduction automatique le « quoi, comment et quel impact ». M. Forcada soulève le problème de la propriété des « data sets » et des algorithmes utilisés par les systèmes de traduction automatique et dans les systèmes hybrides. M. Lacruz Mantecon demande si les machines peuvent être considérées comme auteurs, et comment rémunérer les acteurs participant dans le processus de traduction automatique. Il compare la part humain-machine dans des traducteurs de la première génération utilisant des règles avec ceux utilisant uniquement les données. Les systèmes hybrides complexifient ce partage de contribution. La question des données personnelles est également posée, ainsi que l'avenir du métier de traducteur humain.

La partie suivante est dédiée à la traduction automatique responsable vue par les utilisateurs. Le chapitre 6 est consacré au rôle de la postédition dans le processus de traduction, considéré ici comme écosystème. La traduction automatique constitue une aide au traducteur bilingue ou multilingue qui est situé au centre du système et est soumis aux diverses contraintes comme le facteur temps, exigence du client, et les facteurs sociaux (Krüger). Cependant ce modèle doit être enrichi afin de prendre en compte des aspects éthiques. Trois dilemmes sont mentionnés : le statut de postéditeur, son engagement sur la qualité et la responsabilité liée à la qualité des données utilisées.

Le chapitre suivant présente un écosystème des acteurs (stakeholders) impliqués dans la traduction automatique, qui sont les chercheurs et les développeurs, les

utilisateurs institutionnels et individuels et les postéditeurs. Le degré du risque associé à la traduction dépend de l'environnement et de la culture (pays). Des expérimentations impliquant le japonais, l'anglais, l'allemand et l'espagnol ont été menées, dans l'environnement technique avec le traqueur oculaire et dans l'environnement créatif. À l'issue de ces expérimentations, quatre préconisations éthiques pour les acteurs impliqués ont été formulées afin de préserver l'autorégulation du système : qualité des données, amélioration en continue, responsabilité des traducteurs et des postéditeurs, et responsabilité des chercheurs. Une question reste cependant sans réponse : comment obliger les fournisseurs de traduction automatique à intégrer le retour d'expérience des utilisateurs finaux ?

Les auteurs du chapitre 8 débattent sur l'importance et sur les conditions d'une traduction exacte dans le contexte de la gestion de crise, et préconisent le système hybride et multimodal. Les aspects éthiques abordés ici sont : égalité linguistique et justice sociale face aux situations de crise, parfois en cascade. L'utilité de la « lingua franca » est mentionnée, ainsi que l'importance de « crowdsourcing » multimédia dans la communication vers la population.

Alors que les chapitres précédents exploitent des sujets déjà débattus dans d'autres contextes, et proposent des préconisations qui concernent aussi d'autres utilisations de l'IA, le premier chapitre dans la partie 3 évoque le biais du genre et de l'âge, dans les systèmes commerciaux de traduction automatique Bing et DeepL. Le premier système, Bing, est lié aux spécificités des langues utilisées. Les résultats des expérimentations menées prouvent encore que la qualité des données utilisées pour l'entraînement du système est primordiale.

Le chapitre suivant est consacré à l'empreinte carbone des systèmes neuronaux de traduction automatique. Les auteurs comparent les besoins énergétiques durant la phase d'apprentissage et en exploitation. Des investigations ont été menées pour deux architectures, neuronale classique et transformeur, en utilisant des processeurs standard (CPU) et graphiques (GPU) dans la perspective de « Green AI for NLP ». Un objectif de ce travail est de motiver les chercheurs à élaborer des systèmes de traduction automatique plus efficaces et moins gourmands en énergie en même temps. Les auteurs comparent les impacts environnementaux des traducteurs automatiques en commençant par celui exploitant des règles du transformeur. Les expérimentations démontrent que le choix entre le CPU et le GPU est difficile, car le GPU utilise plus d'énergie mais il est plus performant. Leur choix est de réduire la consommation de l'énergie par la distribution, la parallélisation et la quantisation. Alors que Red AI ne se préoccupe pas de considérations environnementales mais uniquement des performances, le Green AI demande plus d'efforts en termes d'optimisation. L'étude des cas donne les détails sur les différentes approches expérimentées en Irlande et aux Pays-Bas et conclut que la quantisation des modèles donne de meilleurs résultats impacts/performances par rapport aux transformeurs. Cependant, les auteurs ne prennent pas en compte le cycle complet puisque l'impact du data center stockant les données indispensables pour le processus n'est pas intégré.

Le dernier chapitre est consacré aux impacts du traitement de la parole en traduction automatique. Les auteurs soulèvent des questions liées aux utilisations diverses en exploitation possibles de la voix, et particulièrement aux utilisations frauduleuses et à la surveillance (ce n'est pas nouveau, mais cet usage est amplifié grâce à la technologie). Ils mentionnent quelques cas d'utilisation frauduleuse par les grands acteurs technologiques, comme GAFAM et autres. La solution proposée est de généraliser la privatisation de la voix à l'aide de l'anonymisation et l'encodage. La traduction automatique de bout en bout « parole vers parole » et des assistants virtuels devrait être encadrée. Une recherche multidisciplinaire incluant les juristes pourrait faire avancer la sécurité des systèmes multimédias de traduction automatique.

En conclusion, la dernière partie, en particulier l'impact environnemental et les exploitations de la voix, est extrêmement intéressante et devrait faire objet d'un autre ouvrage. L'avenir et la place des traducteurs et des interprètes humains dans cet écosystème méritent aussi une discussion plus importante.