100 MILLIONS DE MOTS TRADUITS
AVEC LE "EU COUNCIL PRESIDENCY TRANSLATOR"
19.11.2020 | Connaissance de la langue et du texte | Technologie linguistique et multilingualité | Sarrebruck

Note : Version française traduite par un système DFKI et minimalement post-éditée de l'interview originale allemande, générée par l'EUCPT.

- Monsieur van Genabith, vous êtes directeur scientifique du DFKI et dirigez depuis 2014 le département Multilingual Language Technologies (MLT) à Sarrebruck. Comment était votre carrière scientifique avant de vous installer à Sarrebruck ?

Le succès du European Council Presidency Translators est une belle occasion pour notre équipe MLT et nos partenaires de DeepL, Tilde et eTranslation ! Je suis très fier des équipes et du travail qu'elles ont accompli en étroite collaboration avec le ministère fédéral des Affaires étrangères ! Je me suis moi-même intéressé depuis très longtemps à la langue et à la technologie, j'ai étudié l'électrotechnique et l'anglistique à l’université RWTH d'Aix-la-Chapelle et j'ai eu beaucoup de chance par la suite : Grâce à une bourse du British Council et, plus tard du Foreign & Commonwealth Office, j'ai pu d'abord faire un MA à l'Université d'Essex puis faire mon doctorat sous la direction de Louisa Sadler. Au début des années 90, j'ai ensuite travaillé comme PostDoc auprès de Hans Kamp à l'Institut de traitement de la langue mécanique (IMS) à Stuttgart. Un bon moment ! Après cela, j'ai passé 17 ans en Irlande à la School of Computing de la Dublin City University, où j'ai passé toute la gamme de cours, de maître de conférences et de professeur associé. À Dublin, j'ai eu de nombreuses libertés et de grands collègues à la DCU, dans les autres universités de Dublin et dans les nombreuses entreprises de haute technologie établies en Irlande (IBM, Microsoft, Symantec), et nous avons pu profiter de ces libertés : J'ai reconstruit le National Center for Language Technology (NCLT) et j'ai été le directeur fondateur du CNGL (Center for Next Generation – aujourd’hui ADAPT et dirigé par Vinny Wade). Grâce à ces travaux, et en particulier au CNGL, nous avons été de plus en plus impliqués au début de la seconde moitié des années 2000-2010 dans des projets internationaux, par exemple de l'UE, dans lesquels l'ancien directeur de notre laboratoire à Sarrebruck, Hans Uszkoreit, était très actif. En 2014, après 17 ans en Irlande, je

***Light post-editing** implies minimal intervention by the post-editor to make the text understandable. Grammar, punctuation and spelling are correct, the translation is complete and accurate in content, but not necessarily idiomatic and fluent.

suis arrivé à Sarrebruck et au DFKI par Hans Uszkoreit, qui avait entre-temps construit le laboratoire frère à Berlin (aujourd'hui SLT, dirigée par Sebastian Möller).

- En plus de votre activité au DFKI, vous avez également une chaire à l'Université de la Sarre. Comment les travaux universitaires et les travaux axés sur les applications se complètent-ils ?
Le plus important dans notre travail sont les collaborateurs et collaboratrices : Grâce à eux, notre travail devient un succès ! Mes collaborateurs de l'Université et du DFKI travaillent ensemble et de manière variée en équipes. Lors de nos réunions hebdomadaires communes, il n'y a pas de différence entre le fait que quelqu'un soit au DFKI ou à l'université. Nous faisons partie du SFB1102 (Information Density and Linguistic Encoding) à l'université, nous avons un projet DFG à l'université à la post-édition multimodale, où nous coopérons avec beaucoup de succès avec l'équipe de professeur Antonio Krüger; je dirige le programme de master européen dans la technologie de langue et de communication (LCT, Erasmus +), qui est présenté par une de mes assistantes de direction au MLT-Lab (DFKI) via un poste universitaire à temps partiel. Tous mes collaborateurs et collaboratrices de la direction du DFKI enseigne dans les quatre groupes de MLT : Machine Translation, Question Answering and Information Extraction, Talking Robots et Data and Resources, donnent des séminaires et forment des étudiants en doctorat, MSc et BSc. De même, de nombreux collaborateurs des équipes MLT sont actifs à l'université. Bien sûr, formellement et financièrement tout est propre séparé en projets. Mais le lien avec l'université est très fort. Le département "Language Science and Technology" de l'Université de la Sarre est l'un des meilleurs en Europe. Au MLTLab du DFKI, nous sommes d'abord forts chercheurs : Nous avons par exemple publié en 2020 plus de 10 documents sur les principales conférences internationales dans notre domaine (ACL, ICML, EMNLP, Coling, IJCAI) dans le domaine de la technologie linguistique, de l'IA et de l'apprentissage mécanique. C'est un grand succès et montre la qualité des équipes. D'autre part, la recherche axée vers les applications du DFKI est une attraction pour les étudiants et les chercheurs de l'université : Où autrement son propre travail, comme par exemple au Conseil de l'UE Presidency Translator, est-il publiquement utilisé pour tous de manière visible de telle sorte que 100 millions de mots soient traduits en 4,5 mois (à ce jour) ? C'est formidable !

- Le EU Council Presidency Translator a continué à promouvoir en Allemagne la visibilité des services de traduction automatique. C'est une prestation collective de plusieurs acteurs, mais vous avez dirigé ce projet. Quand avez-vous commencé à travailler ? Comment avez-vous constitué le consortium ? Et combien de scientifiques ont été impliqués ?
Le European Council Presidency Translator est une solution très européenne qui montre que l'Europe est plus que compétitive au niveau international dans le domaine de la technologie linguistique et de l'IA : Elle repose sur une combinaison d'une expertise exceptionnelle en matière de haute technologie et d'IA en Allemagne (DeepL, DFKI), Lettonie (Tilde) et EC (eTranslation). Un partenariat entre l'industrie (DeepL, Tilde), les pouvoirs publics (EC, eTranslation) et un institut de recherche (DFKI). Le DFKI dirige le projet, le soutien vient du ministère fédéral des Affaires étrangères, qui est l'organisme chef de file de la présidence allemande du Conseil de l'Union européenne. Les compétences des membres du consortium s'y complètent idéalement : Tilde a développé pendant de nombreuses années avec le soutien européen le cadre de base du Presidency Translator, dans lequel les machines de traduction de nombreux fournisseurs sont intégrées, et pilote ses propres machines de traduction. DeepL propose des machines de traduction d'une qualité exceptionnelle pour 8 langues. ETranslation (EC) fournit un service de traduction automatique de base pour toutes les 24 langues officielles de l'UE. En étroite collaboration avec les services de traduction des ministères, le

DFKI a mis au point des systèmes de traduction automatique en allemand, français et espagnol adaptés spécifiquement aux données et aux besoins des Minstéres. Tilde fait cela pour l'anglais, l'italien et le polonais. Au DFKI, Stephan Busemann s'occupe administrativement du Presidency Translator. Je dirige les aspects scientifiques et techniques. Cristina España Bonet, directrice de l'équipe de MT au MLT-Lab et son collaboratrice Jingyi Zhang développent les systèmes. Ils sont soutenus par deux étudiantes, Damyana Gateva et Anastasija Amman, du programme MSc "Language Science and Technology" de l'université. Le DFKI dirige également le travail en amont et médiatique du Presidency Translator. Cela est pris en charge par Eileen Schnur et sa collègue Marlies Thönnissen au sein de l'équipe MLT et soutenu activement par le département de communication d'entreprise de l'DFKI.

- Ils utilisent des réseaux neuronaux artificiels pour la traduction. Pouvez-vous décrire le fonctionnement de votre machine de traduction ?

Ces dernières années, les modèles neuronaux ont permis des avancées quantiques dans la qualité de nombreuses technologies linguistiques et d'autres applications dans l'IA. Notre système utilise des réseaux neuronaux profonds basés sur des modèles de transformateurs. Ces modèles utilisent différents types d'attention et sont largement hautement parallélisables.

- Les réseaux neuronaux artificiels sont formés – testés - avec de très grandes quantités de données linguistiques. D'où viennent ces données de formation et de test et, seulement comme estimation, de combien de mots en cours il s'agit ?

Pour de nombreuses paires de langues, nos meilleures données de formation sont des dizaines de millions de paires de phrases, chaque paire de phrases contenant un ensemble de départ dans une langue et sa traduction dans l'autre langue. De là, les machines apprennent à traduire elles-mêmes. Ces données sont basées sur des traductions déjà faites par l'homme. La machine apprend donc de l'homme. Les données proviennent de collectes de données de l'UE, de l'ELRC (la coordination européenne des ressources linguistiques que nous gérons également au MLT du DFKI) et d'autres sources. En outre, nous travaillons très étroitement avec les équipes de traduction des ministères pour produire des données des ministères sur les machines spécialisées, qui sont particulièrement adaptées aux besoins des ministères. Ceux-ci sont constamment évalués par les traducteurs et traducteurs des ministères, de sorte qu'ils peuvent être continuellement améliorés tout au long du projet.

- Le Presidency Translator a été utilisé intensivement par les utilisateurs au cours des 150 derniers jours. Plus de 100 millions de mots ont été traduits. Quelles étaient les paires de langues les plus demandées ? Et y avait-il peut-être aussi des phrases qui étaient particulièrement fréquentes ?

Contrairement à d'autres offres, le Presidency Translator est sûr et sécurisé, tous les serveurs sont situés dans l'UE, les transmissions sont cryptées, et après une traduction créée, toutes les données sont supprimées immédiatement. Les chiffres montrent que la traduction en un clic du site en langue allemande de la présidence du Conseil est très bien acceptée : Environ 47 % des 100 millions de mots traduits jusqu'à présent sont ainsi réalisés. Les langues cibles privilégiées de la traduction automatique sur le site Internet de la présidence du Conseil sont l'espagnol, l'italien et le portugais (les versions française et anglaise ont été créées manuellement). La moitié un peu plus grande résulte de textes (22 %), de documents (30 %) et de traductions de pages web (2 %) sur la page Translator, et c'est ici que la traduction entre l'allemand et l'anglais est la plus demandée.

- Que disent les traducteurs et traductrices de la nouvelle qualité de la traduction automatique ? Les traducteurs voient-ils les machines comme des concurrents ou comme des outils qui soutiennent leur travail ? Et comment la profession du traducteur change-t-elle ?

Au sein du projet "EU Council Presidency Translator", nous travaillons en étroite collaboration avec les collègues des équipes de traduction des ministères : Ils gèrent la collecte et la mise à disposition de données au sein des ministères afin d'adapter les machines

spéciales aux besoins des ministères. En outre, ils testent et évaluent les machines spéciales et contribuent, grâce à leurs résultats, de manière centrale à l'amélioration des systèmes. Dans le processus de travail de la traduction, les machines sont alors un outil : Avec une bonne qualité de traduction, la machine peut aider à augmenter la productivité d'un traducteur humain. L'image professionnelle du traducteur évolue vers le contrôle de la qualité, l'assurance de la qualité en éditant (rectifiant) des traductions automatiquement créées et en certifiant les traductions et leur qualité. La formation moderne des traducteurs tient compte de ces changements : Le cours de traduction "Translation Science and Technology" à l'Université de la Sarre a une part importante de technologie dans laquelle les futurs traducteurs et traducteurs sont familiarisés avec les technologies linguistiques développées par leurs collègues des cours de linguistique informatique (Science et technologie de la langue) et d'informatique.

- La présidence allemande du Conseil de l'Union européenne prend fin le 31 décembre 2020. Comment le Presidency Translator sera-t-il utilisé par la suite ? Et indépendamment de cela, que ls sont vos plans futurs ?

Le Presidency Translator a été largement bien accepté et a dépassé tous les records précédents de Presidency Translator. Je suis très fier de ce que l'équipe MLT du DFKI a accompli avec ses collègues de DeepL, Tilde et eTranslation ! Il y a un grand intérêt à ce que le Presidency Translator soit utilisé pour d'autres présidences du Conseil. Des discussions sont en cours à ce sujet. De plus, l'industrie s'intéresse beaucoup à la technologie linguistique allemande et européenne : La technologie linguistique et l'IA « made in Europe ». La traduction automatique n'est qu'une des compétences de notre laboratoire MLT : D'autres sont celles du groupe « Question-Answering and Information ExtractionGroup » (en particulier dans le domaine biomédical), du groupe « Talking Robots » (qui se concentre sur les systèmes de dialogue et les robots de la vie) et du groupe « Data and Resources » (qui dirige de grands projets européens comme l’ELRC depuis de nombreuses années). À cela s'ajoute notre laboratoire frère" - SLT (Speech and Language Technology) à Berlin. Les deux laboratoires (MLT à Sarrebruck et SLT à Berlin) coopèrent étroitement et se complètent dans leur expertise.