



HAL
open science

**Tralogy II : Synthèse de la session d'ouverture. Trouver le sens : où sont nos manques et nos besoins respectifs ?
Paris, les 17 et 18 janvier 2013**

Nicolas Froeliger

► **To cite this version:**

Nicolas Froeliger. Tralogy II : Synthèse de la session d'ouverture. Trouver le sens : où sont nos manques et nos besoins respectifs ? Paris, les 17 et 18 janvier 2013. Tralogy II. Trouver le sens : où sont nos manques et nos besoins respectifs ?, Jan 2013, Paris, France. 4p. hal-02497138

HAL Id: hal-02497138

<https://hal.science/hal-02497138>

Submitted on 3 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TRALOGY

Tralogy II : Synthèse de la session d'ouverture

Trouver le sens : où sont nos manques et nos besoins respectifs ? Paris, les 17 et 18 janvier 2013

Nicolas Froeliger

Univ Paris Diderot, laboratoire CLILLAC-Arp, EA 3967
nf@eila.univ-paris-diderot.fr

TRALOGY II - Ouverture de la conférence
Date d'intervention : 17/01/2013

lien video : https://webcast.in2p3.fr/video/translation_and_the_new_digital_commons



La deuxième édition du colloque Tralogy, cette fois avec le titre *Trouver le sens : où sont nos manques et nos besoins respectifs ?* a été ouverte le 17 janvier 2013 par Gilles Sentise, délégué régional du CNRS, pour quelques mots d'accueil au nom de cette institution, hôte de la manifestation. Présidée par Alain Wallon, la session initiale a donné le ton de l'ensemble, avec quatre intervenants invités :

- Daniel Kluvanec (DGT-UE): *MT@EC - Machine Translation at the European Commission*
- Hans Uszkoreit (DFKI, META-NET): *The META-NET Strategic Research Agenda for Multilingual Europe 2020*
- Algirdas Saudargas (Parlement européen): *Translation Strategy for Small Languages*
- Jan Hajic (Institut de linguistique formelle et appliquée, Université Charles, Prague) : *Meaning in Translation : Translators Teaching Machines*

Le président de la session fait tout d'abord observer que l'avancée est notable par rapport à Tralogy I : là où cette première édition, en mars 2011, avait donné le spectacle de trois communautés (*biotraducteurs*, spécialistes du traitement automatique du langage, formateurs) juxtaposées, celle qui s'annonce montre que ces groupes commencent d'interagir – et il faut s'en féliciter dans la mesure où c'est un des objectifs premiers de Tralogy.

La parole est ensuite à Daniel Kluvanec, conseiller du directeur général de la DGT (Direction générale de la traduction, Commission européenne) en matière de traduction automatique. Dans sa présentation du projet MT@EC, celui-ci remarque d'emblée que la traduction, au sein de l'Union européenne, fait partie intégrante du processus législatif, et contribue ainsi à une amélioration de la réglementation et de la régulation dans ce contexte, ce qui a pour effet de faire des traducteurs des gardiens des traités. Sur le plan fonctionnel, ceux-ci sont entourés d'un certain nombre d'outils (ressources terminologiques, documentation, etc.) parmi lesquels, de plus en plus, la traduction automatique (TA). Et de rappeler qu'a eu lieu entre 2010 et 2012 un changement de paradigme à cet égard. Alors que la TA à base de règles dominait depuis 1975, c'est maintenant la TA statistique, objet de MT@EC, qui tient le haut du pavé. D'où une question : cette technologie est-elle suffisamment mature pour un usage professionnel ? Oui et non. Avec le niveau de qualité actuel, il est avant tout possible de sélectionner des textes qui seront ensuite confiés à des traducteurs humains ou qui, en tout cas, ne sont pas destinés à la publication. Améliorer ces outils pour les rendre utilisables par les traducteurs professionnels permettra à la fois de gagner en cohérence sur les gros volumes et de réserver l'attention des agents humains aux composantes les plus intéressantes de l'opération. Cela suppose de réintégrer une part de traduction à base de règles dans la TA statistique et, au passage, d'aller au-delà de systèmes d'évaluation tels que le score BLEU (pour *BiLingual Evaluation Understudy*). Au final, ces questions en appellent d'autres, et d'importance : *qu'est-ce que l'intelligence, et ne faut-il pas la redéfinir ; les compétences humaines sont-elles transférables aux machines ?*, notamment. À ce stade, en tout cas, pointent déjà quelques certitudes : la TA doit s'enrichir d'apports sémantiques et pragmatiques, les solutions passent par l'interdisciplinarité et l'interopérabilité et les traducteurs demeureront les garants de *l'intention* du législateur.

C'est ensuite à Hans Uszkoreit, coordinateur du réseau d'excellence META-NET, qui regroupe une cinquantaine de laboratoires européens, d'intervenir. Sa présentation porte avant tout sur le rôle des technologies langagières aujourd'hui, et est entièrement placée sous le signe du paradoxe. En effet, dit-il, on peut se demander pourquoi ce secteur connaît une telle effervescence alors qu'un grand nombre de problèmes restent à résoudre. La réponse tient à l'augmentation des besoins manifestés par la société et aux progrès remarquables accomplis par ces imparfaites technologies. Deuxième paradoxe, l'ingénierie des langues est partout, mais toujours dans la discrétion. Qui plus est, les avancées majeures ne relèvent plus guère, aujourd'hui, de la recherche : on les trouve désormais au niveau des applications. Dans une phase de concentration, les ténors du secteur ont compris qu'il faut adapter ces technologies – qui sont sans doute l'avenir de l'informatique – aux utilisateurs. Quelle est, dans ce contexte, la place de l'Europe ? Celle-ci a laissé échapper quelques occasions, mais elle peut encore attraper le train. Sa situation

linguistique est différente de celle du reste du monde : c'est pour elle un trésor, un problème et une chance à saisir. Dans le même temps, et c'est un nouveau paradoxe, l'anglais y prédomine, au point de menacer certaines langues « d'extinction numérique ». L'action en faveur d'une société plurilingue est donc déterminante. Celle-ci comporte trois éléments : des programmes appropriés – c'est le rôle de META-NET –, mais aussi des acteurs résolus et des soutiens. Ce sont ces deux derniers aspects qui manquent. Les ressources financières sont là, il faut seulement les dépenser à bon escient, dans l'intérêt des langagiers, des entreprises et des citoyens, et placer ainsi l'Europe devant ses concurrents. META-NET a, pour cela, défini un programme de recherche stratégique, dont une des déclinaisons est la conception d'un « nuage translinguistique » (un système modulaire de traduction via Internet permettant d'accéder à différents services dans sa propre langue à partir d'une plate-forme respectueuse de la confidentialité qui existe actuellement en laboratoire, mais n'est pas encore mûre pour un lancement public). D'autres pistes sont la mise en place d'outils favorisant l'intelligence collective et d'assistants interactifs sensibles aux contextes et aux cultures. Ici aussi, la TA doit aller plus loin dans sa recherche de la qualité, par exemple pour être employée lors de réunions virtuelles ou mise au service de la démocratie numérique et de la prise de décision en facilitant les échanges entre citoyens et experts. La qualité est une condition de l'acceptabilité par les opérateurs humains – et elle passe par la prise en compte de la sémantique. Autre défi technologique à relever, celui des métadonnées, qui ne concernent pas seulement les textes, considérés comme des recueils de pensées complexes, mais aussi le son, l'image, les chiffres et les supports vidéo. Là encore, la solution est en partie sémantique.

Ces questions, comme celles soulevées par le premier intervenant sont, on l'aura compris, politiques. Il était donc logique et nécessaire de donner alors la parole à un représentant de cette sphère, en la personne d'Algirdas Saudargas, aujourd'hui député européen et naguère ministre des Affaires étrangères de la Lituanie lorsque celle-ci a retrouvé son indépendance, en 1990. Son pays d'origine – et n'oublions pas que les pays baltes sont les invités de cette deuxième édition de Tralogy – est représentatif de ceux qui sont menacés, comme le disait Hans Uszkoreit, « d'extinction numérique ». Ce qui semble être le cas, selon un livre blanc de META-NET, de 21 langues officielles européennes sur 30 (<http://www.meta-net.eu/whitepapers/overview-fr>). Question existentielle qui appelle des réponses politiques, seules à même d'emporter une adhésion populaire. L'Histoire de la Lituanie est exemplaire à cet égard : par le passé, la langue de ce pays a souffert d'un manque d'accès aux technologies, et en particulier à l'imprimerie, et l'adoption de l'alphabet cyrillique a été un pis-aller pour surmonter cet obstacle... Aujourd'hui, la problématique de l'accès est tout autre : l'information est à portée de main pour chacun, mais quelle information, avec quel degré de complexité dans sa compréhension ? Et de faire référence à l'essai de Nicolas Carr *The Shallows—What the Internet is Dosing to Our Brains* (*Internet rend-il bête*, en français, Robert Laffont, 2011). Nous avons l'accès au savoir, mais pas le savoir lui-même, notamment parce que celui-ci est formulé, pour une large part, en anglais, et que le projet de diffuser la totalité des connaissances humaines dans une langue unique a échoué. Pour une connaissance véritable, approfondie, il faut réhabiliter les langues dans leur diversité. Voilà pourquoi les technologies langagières sont indispensables. Celles-ci doivent donc être réinvesties dans une approche globale. Comme l'a laissé entendre l'orateur précédent, les langues (comme d'ailleurs les traducteurs) sont partout, mais elles ne sont pas mises en avant. Il importe de le faire, à travers des projets ambitieux et visibles, qui valorisent le cerveau humain – et, là encore, la sémantique. De par sa diversité linguistique, l'Europe, là encore, est bien placée pour cela.

Le dernier intervenant de la session était Jan Hajic, ancien directeur de l'Institut de linguistique formelle et appliquée à l'Université Charles de Prague. En introduction, celui-ci revient sur l'évolution de la traduction automatique. Après une TA à base de règles, celles-ci étant créées par des humains et fonctionnant selon le schéma analyse-transfert-synthèse, on est maintenant passé à la TA statistique, fondée sur l'apprentissage par la machine, à partir de classifications (humaines) de données. Ce principe, adopté par exemple par la plate-forme Moses, fournit, nous dit l'orateur, de bons résultats à l'échelle des mots et des expressions. L'avenir, néanmoins, appartient aux solutions hybrides, c'est-à-dire reposant sur les méthodes statistiques, mais intégrant une analyse linguistique approfondie sous forme d'arborescences. C'est l'objet du

programme EUROMATRIX. Il s'agit notamment de reproduire informatiquement les schémas de classification des données mis en œuvre par les opérateurs humains, en utilisant aussi bien des traductions alignées que des textes monolingues (qui renseignent notamment sur l'usage ou la phraséologie). L'intervenant reprend ici le triangle de Vauquois (analyse de la langue source à gauche, transfert à l'horizontale, production d'un texte d'arrivée à droite, avec comme pivot une représentation structurale universelle), pour proposer de substituer à la représentation linguistique de surface (très différente dans chaque langue) une représentation sémantique qui, elle, sera très proche dans la langue de départ comme dans celle d'arrivée. Le tout donc, en procédant par arborescences. Pour s'affranchir des insuffisances de l'analyse linguistique, la machine doit ainsi apprendre non seulement la langue, mais aussi le sens. Ce qui suppose de l'alimenter par des traductions de qualité, avec à terme une perspective d'inversion des vecteurs : ce système a lui aussi vocation à produire des traductions performantes. Conclusion : pour avancer dans ce domaine, il faut des données et donc des traducteurs, en même temps que des capacités d'analyse affinées.

S'agissant d'une session introductive, cette partie du colloque n'a pas donné lieu à débat. Il importe simplement d'observer que ces quatre intervenants ont su se placer dans la droite ligne des intentions de cette manifestation, sous le maître mot partagé d'hybridation – et que leurs observations sont remarquablement concordantes. Un très grand nombre des idées formulées à cette occasion se verront ainsi reprises et développées par d'autres intervenants lors des présentations qui suivront. Riche et prometteur départ, donc.