



HAL
open science

Rapport sur la septième session de la conférence Tralogy I: La disponibilité des ressources

Jean-François Nominé, Martine Garnier-Rizet

► To cite this version:

Jean-François Nominé, Martine Garnier-Rizet. Rapport sur la septième session de la conférence Tralogy I: La disponibilité des ressources. Tralogy I. Métiers et technologies de la traduction: quelles convergences pour l'avenir?, Mar 2011, Paris, France. 6 p. hal-02496072

HAL Id: hal-02496072

<https://hal.science/hal-02496072>

Submitted on 2 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.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0
International License

TRALOGY

Rapport sur la septième session de la conférence Tralogy I : La disponibilité des ressources

Jean-François Nominé
Inist-CNRS, France

Martine Garnier-Rizet
IMMI-CNRS & Vecsys, France

TRALOGY I - Session 7
Date d'intervention : 04/03/2011



La session intitulée "Disponibilité des ressources" constitue la septième session de la Conférence Tralogy'11. Jean-François Nominé introduit le sujet en rappelant l'importance fondamentale que revêt la question de la disponibilité des ressources linguistiques, qu'il s'agisse du point de vue du traducteur humain ou de celui des chercheurs et ingénieurs du domaine de la traduction automatique. Cette notion de disponibilité recouvre plusieurs aspects qui ont été évoqués à plusieurs reprises déjà lors de la conférence. Sous quelles conditions ces ressources doivent-elles être mises à disposition ? Quels sont les droits d'utilisation de ces ressources ? Pour quels usages ?

Ces questions posées, Jean-François Nominé présente le premier orateur Philippe Lacour dont l'intervention s'intitule « Translation and the New Digital Commons », préparée en collaboration avec Any Freitas (*European University Institute*). Chercheur en philosophie du langage à l'École Normale Supérieure mais également fondateur il y a cinq ans d'une association qui porte le projet Traduxio, Philippe Lacour introduit son exposé en soulevant la question de la diversité linguistique. Comment peut-on concevoir la diversité linguistique ? Doit-elle être considérée comme une limite à l'échange - on utilise souvent le terme de barrière linguistique- ou plutôt comme une richesse à entretenir de façon durable ? Une jolie métaphore suggère d'envisager deux langues comme deux vallons d'un paysage, avec le choix pour aller de l'un à l'autre soit d'aller vite en empruntant un chemin rapide soit de se laisser mener plutôt par les chemins de traverses en prenant le temps de découvrir toutes les richesses offertes. Philippe Lacour insiste sur la non exclusion de ces approches. Sans remettre en question l'apport des nouvelles technologies, il a choisi dans le projet Traduxio de s'intéresser à la deuxième approche, celle de la diversité.

Après avoir rappelé qu'un des premiers phénomènes multilingues du web est l'encyclopédie collaborative Wikipédia, Philippe Lacour pose la question suivante, initialement énoncée par Salah Basalamah de l'Université d'Ottawa : y a-t-il un droit de traduire, qui serait coextensif à la mondialisation et participerait à la propagation des savoirs ? Et pour tenter de répondre, Philippe Lacour choisit de décrire un cas concret, le projet Traduxio, initié il y a plusieurs années. Ainsi que l'indique la page d'accueil au travers des trois mots « Compare », « Interpret », « Share », Traduxio constitue un environnement de traduction collaboratif qui permet de comparer d'interpréter et de partager un certain nombre de traductions. Cet atelier collaboratif n'est pas forcément un projet éditorial. Il est gratuit, de façon durable, sous licence GNU GPL3. Bien qu'il soit prévu pour la traduction littéraire ou les textes culturels parce que la difficulté sémantique que l'on cherche à aborder est immédiate dans ces cas-là, d'autres usages sont possibles, qui seront imaginés sans doute par l'auditoire issu pour partie des métiers de la traduction. Philippe Lacour insiste sur le fait qu'il ne s'agit pas de traduction automatique ou d'un quelconque traitement automatique du langage mais d'un environnement un peu particulier d'aide à la traduction qui n'est pas au sens strict une mémoire de traduction. Les différents acteurs sont ensuite présentés - l'association (loi 1901), à Paris et le laboratoire Tech-Cico de l'Université de Troyes, spécialisé dans les technologies d'Interprétation Assistée par Ordinateur, les technologies multi-points de vue, les technologies de conflits d'interprétation - ainsi que les partenaires qui ont financé le projet - l'UNESCO, l'Organisation Internationale de la Francophonie (OIF) et la Délégation Générale à la Langue Française et aux Langues de France (DGLFLF). Le projet a incubé au centre Marc Bloch de Berlin, un centre en sciences sociales où chacun parle sa langue, le français ou l'allemand.

Philippe Lacour souligne l'originalité du projet Traduxio qui consiste à refuser l'alternative, souvent exposée, entre la traduction de masse et le savoir-faire artisanal inhérent à la traduction des textes difficiles. Résister à cette alternative, c'est montrer que la technologie peut être mise au service de fonctionnalités proprement humaines, qui font la valeur de la traduction humaine. Ceci repose sur une conception légèrement différente de l'intelligence artificielle : « passer des machines qui pensent aux machines qui donnent à penser ». Plutôt que de demander à la machine de traduire à la place de l'humain, il s'agit ici de faire appel aux mécanismes de la suggestion et demander comment une personne a traduit auparavant, éventuellement dans une langue différente.

Et là réside une des originalités du projet Traduxio, qui est de ne pas considérer une configuration en étoile, par couple de langue, en utilisant notamment une langue pivot. Ainsi Google utilise-t-il l'anglais comme langue pivot que ce soit pour traduire du japonais au coréen, ou du flamand à l'afrikaans, ce qui est plus discutable si l'on considère la proximité de ces deux dernières langues. Il s'agit ici au contraire de privilégier une configuration en série, avec une traduction systématique d'une langue vers plusieurs. Philippe Lacour illustre son propos à partir d'un exemple de traduction d'Hamlet en français, espagnol et allemand. Il souligne aussi l'importance de la prise en compte de l'histoire de la langue.

La seconde originalité est le recours aux recherches dans des documents déjà traduits, ce qui s'apparente aux mémoires de traduction, à la différence que le traducteur a pu introduire au préalable ses propres catégories d'étiquetage collaboratif et sa propre segmentation (qui n'est pas limitée à la phrase) ; on parle ici plutôt de bases documentaires. Philippe Lacour présente ensuite différentes fonctionnalités de l'outil, relatives à la recherche dans les bases documentaires.

La troisième et dernière originalité de Traduxio, bien que plus courante, notamment dans Wikipedia, repose sur la notion d'historique des révisions ainsi que le choix de rendre publique ou non sa traduction.

L'objectif déclaré d'un tel projet est de valoriser les compétences propres du traducteur humain, en s'inscrivant dans le mouvement de la culture « libre ». Chaque utilisateur, traducteur, enseignant, chercheur, est invité à créer sa base documentaire, en bénéficiant de la souplesse offerte par l'outil.

Ceci amène au dernier point évoqué par Philippe Lacour, qui revêt une importance particulière, celui des aspects juridiques. Après avoir rappelé les caractéristiques de la licence *Creative Commons*, qui notamment exclut les œuvres dérivées, il présente la réflexion qui a été menée au sein de l'association pour adapter ce type de licence au domaine de la traduction - la traduction, rappelons-le, étant justement considérée comme une œuvre dérivée. Philippe Lacour précise qu'il y a peut-être un moyen d'aménager les droits d'auteur pour ménager les droits des traducteurs, et plus précisément toute la chaîne des traducteurs lorsque ceux-ci travaillent dans une démarche collaborative. Pour illustrer son propos, il évoque la démarche du Fonds Paul Ricoeur, qui souhaite diffuser la pensée du philosophe notamment à travers la traduction de ses œuvres, dont le Fonds détient les droits. Philippe Lacour soumet alors la proposition d'un aménagement de la licence *Creative Commons* (CC+) dans lequel les traductions échapperaient à l'exclusion des œuvres dérivées à condition que soient respectées les conditions suivantes ; traduire tout le texte, ne traduire que le texte, traduire à partir de l'original.

Philippe Lacour conclut son intervention en indiquant que l'association est à la recherche de partenaires, que ce soit pour tester l'environnement de traduction ponctuellement, ou de façon plus suivie. Il s'adresse à la communauté des traducteurs en insistant sur la démarche du projet Traduxio, qui, bien loin de vouloir s'appropriier le travail d'autrui, est dans une démarche de préservation et d'enrichissement collaboratif. Il mentionne enfin deux articles à l'intention des littéraires pour le premier et des informaticiens pour le second (cf. dernière diapositive de la présentation).

Jean-François Nominé remercie Philippe Lacour d'avoir évoqué les aspects des droits juridiques, qui dit-il ne manqueront pas de susciter des questions dans l'assemblée. Il introduit ensuite le deuxième intervenant, François Brown de Colstoun, Président de la société *Lingua et Machina*. Il le remercie d'avoir accepté « au pied levé » de représenter Andrew Joscelyne, souffrant, et de présenter l'association TAUS. Il ajoute qu'à la suite de ce premier exposé, François Brown de Colstoun poursuivra par une courte présentation des activités de *Lingua et Machina*.

François Brown de Colstoun présente l'association TAUS qui, ajoute-t-il, constitue presque un syndicat. Les membres de TAUS sont les grands groupes, tel IBM par exemple, qui effectuent de

nombreuses traductions automatiques mais aussi des entreprises, comme la société Systran, qui développent et commercialisent des systèmes de traduction automatique. L'objectif de TAUS est de mutualiser des mémoires de traduction, essentiellement dans une démarche d'alimenter des systèmes de traduction automatique statistiques. L'objectif annoncé par TAUS est de convaincre les traducteurs de partager leurs mémoires de traduction arguant du fait que chacun bénéficie de ces échanges. La couverture des domaines traités est très large et le nombre de mots très élevé. Le principe est le suivant, les membres ou non-membres, peuvent poster et télécharger des mémoires de traduction. Andrew Joscelyne, qui ne pouvait être là, a souhaité préciser qu'il y aurait bientôt un nombre bien plus important de mémoires de traduction mises à disposition gratuitement. François Brown de Colstoun présente ensuite une diapositive qui résume les différents services proposés par TAUS et de quelles façons les membres utilisent ces services aujourd'hui. La démarche est de faire partager le plus grand nombre de mémoires au plus grand nombre et de limiter les mémoires payantes aux objectifs industriels. On peut accéder également à des outils terminologiques et à des outils de traduction automatique selon que l'on est membre ou pas.

François Brown de Colstoun souligne la qualité de transparence dont fait preuve TAUS vis à vis de sa feuille de route depuis la création en juillet 2008. En ce sens, dit-il, TAUS se comporte comme un syndicat professionnel, qui publie régulièrement plusieurs articles de fond, le dernier portant sur la multiplication des usages de Moses, par exemple. Une question importante qu'il soulève est celle de la pertinence d'une entité comme TAUS dans le mouvement vers le « Cloud », évoqué plus tôt dans l'après-midi. Sans vouloir répondre à la place d'Andrew Joscelyne, François Brown de Colstoun considère que TAUS est sans doute l'organisme non-professionnel le plus visible et le plus actif aujourd'hui dans le domaine.

Il passe ensuite à la présentation de la société *Lingua et Machina*, « l'écrit multilingue dans l'entreprise », conçue en trois parties. *Lingua et Machina* est un éditeur de logiciel, issu du monde des mémoires de traduction, créé il y a huit ans par Emmanuel Planas, et qui collabore avec de nombreux laboratoires. Bien que ce soit, dit-il, une « tarte à la crème », François Brown de Colstoun souligne le changement de paradigme proposé par *Lingua et Machina*. Il s'agit de placer les collaborateurs industriels au cœur du processus, pour qu'ils corrigent et alimentent les mémoires de traduction de l'entreprise. Pour illustrer son propos, François Brown de Colstoun prend l'exemple d'un chef de projet de la société TOTAL confronté à la traduction d'un document très technique de quinze pages, que le traducteur seul ne peut produire. La solution que propose *Lingua et Machina* est de construire une « mémoire d'entreprise » qui s'enrichit des traductions produites au fil du temps. Il s'agit ainsi d'une encyclopédie maison, qui constitue un référentiel commun, qui permet à tout collaborateur de lire, rédiger, corriger et traduire dans une langue étrangère tout type de document, simple courrier électronique ou article de fond.

François Brown de Colstoun passe ensuite du français à l'anglais pour évoquer les principes fondamentaux de la société et les principes opérationnels tels que la gratuité pour les « *freelancers* », les académiques et les étudiants. Il conclut sa présentation en soulignant l'importance de respecter la confidentialité des ressources terminologiques produites pour leurs clients qui sont le plus souvent des grands groupes internationaux.

Khalid Choukri, secrétaire général de ELRA, est le troisième et dernier intervenant de cette session. Il souhaite axer sa présentation sur les activités « données » et « évaluation » de l'association avec un focus sur la traduction automatique. ELRA est une association à but non lucratif, créée en 1995, sur l'incitation de la Commission Européenne qui avait fait le constat qu'une partie du financement de chaque projet était systématiquement consacrée à la production de ressources linguistiques -corpus, lexique, base de données terminologiques, etc. – sans connaissance ou sans accès possible à d'éventuelles ressources déjà produites... et financées lors de projets antérieurs. Ainsi, la mission d'ELRA depuis 15 ans est d'archiver ces données et de les rendre disponibles, en premier lieu aux acteurs du domaine de l'ingénierie de la langue, chercheurs et industriels, qui visent à développer des technologies. La mise à disposition des ressources linguistiques implique de traiter les aspects techniques (support physique, numérisation), légaux (propriété intellectuelle, droit d'exploitation) et commerciaux qui y sont liés.

De plus, dans le cadre de sa mission de dissémination de l'information, ELRA organise tous les 2 ans, depuis 1998 la Conférence LREC (*Language Resources and Evaluation Conference*) qui attire un public international de plus de mille participants.

Khalid Choukri poursuit sa présentation en présentant les différents services offerts aux membres de l'association dont le Catalogue ELRA qui recense plus d'un millier de ressources. Il revient ensuite sur le rôle d'intermédiaire joué par ELRA entre les producteurs et les utilisateurs de données (recherche des données existantes, négociations juridiques, validation de la qualité des ressources etc.) en ajoutant qu'il arrive également qu'ELRA soit lui-même producteur de ressources, en général suite à une commande particulière.

Au fil du temps, et naturellement de par la neutralité de sa position, ELRA a également été amenée à jouer un rôle d'évaluateur des technologies produites par les acteurs du domaine, soit en mettant à disposition des plates-formes d'évaluation de systèmes soit en organisant des campagnes d'évaluation au niveau national ou international (traduction automatique, recherche d'information, résumé automatique, reconnaissance de la parole etc.). Il mentionne également le portail « *HLT Evaluation* », qui recense les différentes initiatives au niveau international.

Puis Khalid Choukri développe son propos sur l'importance de l'approche méthodologique de l'évaluation en insistant sur le choix des métriques utilisées, dans le domaine de la traduction automatique de la parole (« *Speech to Speech Translation* ») notamment, dans le cadre du projet TC-STAR. Il souligne le fait que selon la tâche, les performances d'un système de traduction automatique peuvent être encore éloignées de celles d'un traducteur humain qui ne se contente pas de traduire mais « interprète » voire « édulcore » parfois les propos d'autrui qu'il doit traduire.

Il conclut en présentant la matrice des ressources bilingues disponibles, issue du projet Euromatrix¹. Il rappelle qu'un système de traduction automatique repose soit sur des règles et dans ce cas nécessite des années de développement, soit sur des modèles statistiques et exige alors des données d'apprentissage en grande quantité. La matrice présentée parle d'elle-même et montre le fossé qu'il reste à combler entre les langues très dotées et celles pour lesquelles trop peu de ressources sont aujourd'hui disponibles.

Jean-François Nominé donne la parole à la salle après avoir remercié les intervenants pour leurs exposés qui, dit-il, sont assez différents dans leur nature et leur objet et ne manqueront pas de susciter des questions.

Edouard Geoffrois (DGA) entame le débat en posant la question suivante qui s'adresse plus particulièrement à Philippe Lacour. Peut-on voir une analogie conceptuelle entre la traduction d'un texte, que Philippe Lacour propose de ne pas considérer comme une œuvre dérivée, et la transcription d'un document audio par exemple ? Philippe Lacour répond par l'affirmative en indiquant qu'il s'agit de protéger les corpus et les annotations et en même temps de ne pas exclure la valorisation commerciale. D'après lui, tout reste encore à inventer, dans la direction de *Creative Commons* qui aujourd'hui, avec d'autres, propose de nouveaux types de licences.

A son tour, Claudio Menezes de l'Université de Brasilia au Brésil s'adresse à Philippe Lacour pour lui demander s'il pense que l'outil Traduxio peut favoriser la présence d'autres langues peu répandues dans le cyberspace. Il enchaîne par une seconde question liée à la précédente : Comment Traduxio opère-t-il en dehors de l'espace de l'Union Européenne ? Des partenariats avec d'autres organismes internationaux soutenant le multilinguisme sont-ils envisagés, par exemple en Amérique latine ?

Philippe Lacour souhaite répondre aux deux questions en même temps car il y voit un lien fort, celui de la défense des langues dites minoritaires, souvent peu présentes car elles ne répondent pas à la norme de rentabilité commerciale qui est celle en usage aujourd'hui. Il mentionne le

(1) <http://www.euromatrix.net/>

réseau MERCATOR², par exemple, qui a manifesté son intérêt, bien que le projet Traduxio était à ce moment-là dans une version non encore aboutie. Philippe Lacour conclut en indiquant que l'association a effectivement quelques contacts en dehors de l'Union européenne dans des espaces culturels où la question de la traduction est envisagée comme une priorité politique, en particulier en Corée et en Inde, et qu'elle est toujours ouverte à d'autres échanges et pourquoi pas avec l'Amérique latine.

John Moran prend la parole en évoquant son métier premier qui était d'enseigner à des traducteurs. Il s'adresse à Philippe Lacour en indiquant combien, d'après lui, les étudiants en traduction, ainsi que les enseignants, apprécieraient de pouvoir utiliser un outil tel que Traduxio.

La clôture de la session approche et Jean-François Nominé donne la parole à une dernière personne, François Abraham, Président de l'Ordre des traducteurs, terminologues et interprètes agréés du Québec (OTTIAQ). François Abraham s'appuie sur la présentation de François Brown de Colstoun et regrette que la relation client-traducteur n'intervienne qu'à la fin du travail, alors que des échanges préalables éviteraient les problèmes aigus rencontrés par le groupe TOTAL, mentionnés par *Lingua et Machina*. Cela mène, dit-il, à devoir faire du travail « de rattrapage ». Il souligne enfin que le choix de TOTAL d'externaliser son service de traduction en est sans doute la cause première.

François Brown de Colstoun répond en évoquant la prééminence des services achats des grands groupes et leur volonté de réduire les coûts liés aux travaux de traduction, ce qui, au fil des années, a eu pour conséquence une baisse de qualité certaine, y compris au sein d'agences de renom. Le mot de la fin sera néanmoins plus positif car, d'après lui, on observe depuis peu une prise de conscience de la part des entreprises qui laisse espérer un retour vers davantage de moyens pour satisfaire l'exigence de qualité.

(2) Le réseau MERCATOR est un réseau d'information, de documentation et de recherche en faveur des langues et cultures régionales ou minoritaires