



HAL
open science

Tralogy II. Compte rendu de la session 4 - Terminologie et lexicologie

Hanna Martikainen

► **To cite this version:**

Hanna Martikainen. Tralogy II. Compte rendu de la session 4 - Terminologie et lexicologie. Tralogy II. Trouver le sens : où sont nos manques et nos besoins respectifs ?, Jan 2013, Paris, France. 3p. hal-02497481

HAL Id: hal-02497481

<https://hal.science/hal-02497481v1>

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 II. Compte rendu de la session 4 - Terminologie et lexicologie

Hanna Martikainen

► **To cite this version:**

Hanna Martikainen. Tralogy II. Compte rendu de la session 4 - Terminologie et lexicologie. Tralogy II. Trouver le sens : où sont nos manques et nos besoins respectifs?, Jan 2013, Paris, France. 3p. hal-02497481

HAL Id: hal-02497481

<https://hal.archives-ouvertes.fr/hal-02497481>

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

Compte rendu de la session 4 – Terminologie et lexicologie

Hanna Martikainen

TRALOGY II - Session 4
Date d'intervention : 18/01/2013



Comme l'a fait valoir la présidente de session, Sandrine Peraldi, elle-même engagée de longue date dans l'activité terminologique sous ses différents aspects, il est difficile aujourd'hui d'évoquer ces deux disciplines intimement liées à la traductologie que sont la terminologie et la lexicologie sans parler des outils informatiques qui vont avec, nécessaires notamment pour gérer les volumes considérables de données produits par ces activités. Cette session 4 a donc été l'occasion de partir à la découverte des outils permettant, dans la thématique du colloque Tralogy II, de capter le sens de ces unités terminologiques au cœur de l'activité traduisante, un sens de plus en plus complexe dans le contexte mondialisé qu'est le nôtre. Les différentes interventions de cette session, consacrées respectivement à deux applications dictionnairiques et à deux plateformes de services terminologiques, ont permis de prendre la mesure des incessants progrès technologiques permettant de faire avancer aujourd'hui les connaissances terminologiques et lexicologiques au service du traducteur.

Ingénieure de recherche au laboratoire d'analyse et de traitement informatique de la langue française (ATILF), unité de recherche mixte affiliée au CNRS, Pascale Bernard nous a présenté *Le trésor de la langue française informatisé* (TLFi) à l'usage des traducteurs. Véritable base de connaissances puissante, cet outil informatique permet de puiser dans toute la richesse du *Trésor de la Langue Française* (TLF), de ses 16 volumes rédigés sur presque 30 ans, pour mener des études sémantiques extrêmement précises. Ce degré de finesse est rendu possible par le balisage XML intégral du contenu, ce qui en rend tous les éléments interrogeables. Le processus d'informatisation a abouti à une hiérarchie des données balisée au niveau le plus fin : c'est, en effet, le seul dictionnaire électronique à avoir balisé non seulement les objets, mais aussi leurs dépendances.

Le portail TLFi, en accès libre et gratuit, offre un éventail de ressources lexicales ayant pour point commun une grande liberté dans les paramètres de recherche, dans l'optique de rendre accessible à tout public ce dictionnaire conçu à l'origine pour des chercheurs. Les fonctions pensées pour tout public vont de la liberté de la forme de la recherche (correcteur d'erreurs, recherche par écriture phonétique) aux possibilités de recherche onomasiologique – pour retrouver, en partant de son sens, ce mot irritant que l'on a juste sur le bout de la langue. Parmi les fonctions de recherche avancée particulièrement intéressantes pour l'utilisateur professionnel, citons la recherche par niveau de langue ou indicateur d'emploi, par code grammatical ou discipline, ou encore par synonymie. La liberté de combinatoire des paramètres de recherche permet d'affiner les requêtes jusqu'à atteindre un degré de précision extrême pour explorer les trésors du TLF dans cette quête du sens au cœur de l'édition 2013 de Tralogy.

Dans la conférence suivante, intitulée *Contemporary dictionary application : tips and tricks of getting it right*, Linas Valiukas de l'Institut de la langue lituanienne nous a fait bénéficier de son expérience en tant que développeur de logiciels pour nous révéler en quelque sorte l'envers du décor, c'est-à-dire l'environnement technologique des applications dictionnairiques. En matière d'astuces, il s'agit en fait de bonnes pratiques de l'ingénierie logicielle ayant fait leurs preuves dans le domaine du développement logiciel mais souvent méconnues des linguistes impliqués dans de tels projets. Linas Valiukas a notamment plaidé pour un modèle de développement itératif sans exigences fixes, couplé à une technique de rétroplanification de projet appelé *timeboxing*. Les seuls impératifs fixes du projet concernent les délais de production, tandis que les livrables, flexibles, sont mis au point par itérations successives jusqu'à l'échéance finale. Pour garder trace des modifications apportées au produit tout au long de son développement, il faut en outre impérativement assurer la gestion des versions et le suivi des problèmes à l'aide d'outils adaptés. Enfin, l'orientation utilisateur sous-tendant le développement d'une telle application se traduit, entre autres, par le choix d'un environnement de développement multiplateforme pour n'imposer aucun système d'exploitation par défaut, ainsi que par celui d'une livraison logicielle par internet.

Le quatuor pluridisciplinaire représenté par Olga Vodopiyanova regroupe deux des cinq partenaires européens du projet TaaS (*Terminology as a Service*). Ce projet, financé par le septième programme-cadre de l'Union européenne pour la recherche et le développement (FP7), vise la création d'un environnement cloud complet pour le travail terminologique, autant pour

l'utilisateur humain que pour la traduction automatique. La future plateforme TaaS englobera toute la chaîne de production des données terminologiques multilingues – l'objectif étant de proposer à terme toutes les langues officielles de l'UE – dans une variété de domaines. La force de ce projet repose sur une enquête exhaustive conduite auprès d'utilisateurs potentiels (traducteurs, interprètes, rédacteurs techniques, etc.), dont les résultats permettront de répondre au mieux aux besoins réels des utilisateurs en leur offrant un accès instantané aux termes les plus récents, ainsi que des solutions pour le partage de ces données.

L'avantage principal du portail TaaS pour l'utilisateur traducteur est la réduction de la charge de travail terminologique : il pourra procéder à l'extraction automatique de candidats termes à partir de documents qu'il aura fournis, ainsi qu'à la recherche automatique d'équivalents de traduction dans les langues de son choix dans différentes bases de données terminologiques et sur le web. Des solutions seront offertes pour stocker ces données terminologiques – récoltées automatiquement et nettoyées par l'utilisateur – dans un répertoire partagé sur le cloud, et pour les partager sur des plateformes existantes en accès libre, comme EuroTermBank ou IATE, afin de faciliter leur réutilisation dans d'autres applications ou dans la traduction automatique. L'objectif est d'instaurer une sorte de cercle vertueux et exponentiel d'acquisition, de traitement et de réutilisation de données terminologiques combinant des outils performants pour la collecte automatique de données et l'apport de l'utilisateur humain dans le traitement de celles-ci, le volume de données étant appelé à croître en même temps que la masse d'utilisateurs.

La plateforme de services terminologiques issue du projet TTC (*Terminology extraction, Translation tools and comparable Corpora*) que nous a présentée Helena Blancafort, déjà présente lors de la première édition de Tralogy en 2011, fait figure de digne prédécesseur de la future plateforme TaaS. Si les deux projets ont aussi en commun certains partenaires, le projet TTC, aujourd'hui abouti, se distingue par l'apport des corpus comparables au travail terminologique ainsi que par sa visée TA/TAO. La plateforme TTC comprend trois modules, utilisables en chaîne pour construire des terminologies bilingues étape par étape à partir de zéro, ou indépendamment les uns des autres pour des tâches spécifiques. En premier lieu, un crawler thématique compile des corpus monolingues sur internet à partir de mots-clefs du domaine définis par l'utilisateur. Un outil d'extraction repère ensuite les candidats-termes dans ces corpus – termes simples et compliqués, variantes terminologiques – avant de passer à l'étape d'alignement résultant en des terminologies bilingues avec plusieurs candidats de traduction pour un terme donné. Enfin, l'utilisateur peut établir une connexion avec EuroTermBank, qui gère les terminologies issues du projet, pour éditer les données et partager ses terminologies.

Pour un public de traducteurs professionnels, les atouts sont nombreux : par exemple, la compilation de corpus permet de se familiariser rapidement avec un nouveau domaine, tandis que l'outil d'extraction peut être utilisé pour découvrir la terminologie du domaine aussi bien que pour extraire la terminologie d'un corpus fourni par le client. Les retours recueillis lors d'ateliers utilisateurs ont d'ailleurs été très positifs. Les utilisateurs apprécient particulièrement d'avoir à disposition sur une même plateforme toute la chaîne de travail terminologique, ainsi que la liberté offerte par l'architecture modulaire et les possibilités de paramétrage des outils pour différents profils et usages. Le point le plus délicat est l'utilisation du web comme ressource pour le travail terminologique qui, outre qu'elle entraîne une dépendance des moteurs de recherche limités ou payants, implique toujours la possibilité d'erreurs devant être corrigées par l'utilisateur.

Comme souligné par la présidente de session, le fil conducteur sous-tendant toutes les interventions de cette session 4 était clairement cette volonté de replacer l'utilisateur au centre du processus de développement des outils technologiques qui se manifeste, par exemple, dans les enquêtes de terrain menées auprès des utilisateurs pour connaître leurs besoins réels, dans l'offre de services à la demande, dans le souci d'ergonomie des interfaces, ou encore dans la personnalisation de l'expérience utilisateur. Cette orientation en fait aujourd'hui des outils véritablement au service de l'utilisateur – homme ou machine – pour l'accompagner dans sa quête du sens. Quête que risquent, hélas, d'entraver notamment les limites imposées au partage des données découlant des questions de propriété intellectuelle.