



ILCEA

Revue de l'Institut des langues et cultures
d'Europe, Amérique, Afrique, Asie et Australie

19 | 2014

**Acceptabilité et transgression en langues et cultures
de spécialité**

Du contexte didactique aux pratiques professionnelles : proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée

*Designing a Multiple-Criteria Specialized Translation Quality Assessment Grid
for Both Academic and Professional Purposes*

**Daniel Toudic, Katell Hernandez Morin, Fabienne Moreau, Franck Barbin
et Gaëlle Phuez**



Édition électronique

URL : <http://journals.openedition.org/ilcea/2517>

DOI : 10.4000/ilcea.2517

ISSN : 2101-0609

Éditeur

UGA Éditions/Université Grenoble Alpes

Édition imprimée

ISBN : 978-2-84310-280-6

ISSN : 1639-6073

Référence électronique

Daniel Toudic, Katell Hernandez Morin, Fabienne Moreau, Franck Barbin et Gaëlle Phuez, « Du contexte didactique aux pratiques professionnelles : proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée », *ILCEA* [En ligne], 19 | 2014, mis en ligne le 27 juin 2014, consulté le 19 avril 2019. URL : <http://journals.openedition.org/ilcea/2517> ; DOI : 10.4000/ilcea.2517

Ce document a été généré automatiquement le 19 avril 2019.

© ILCEA

Du contexte didactique aux pratiques professionnelles : proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée

Designing a Multiple-Criteria Specialized Translation Quality Assessment Grid for Both Academic and Professional Purposes

Daniel Toudic, Katell Hernandez Morin, Fabienne Moreau, Franck Barbin et Gaëlle Phuez

- 1 Comme l'admet (avec une certaine ingénuité) un récent article sur l'apport de la post-édition en traduction automatique, « Translation quality assessment is far from a solved problem » (Green *et al.*, 2013, p. 442), la notion de qualité en traduction a longtemps été la préoccupation des seuls théologiens, traducteurs littéraires ou pédagogues. Elle est alors centrée sur la notion de fidélité (ou d'infidélité) au texte source, pour des raisons religieuses, philosophiques, culturelles ou didactiques. Dans les années 1980, le développement de la traduction spécialisée ou « pragmatique », répondant à des besoins commerciaux, industriels, juridiques, médicaux, sociaux ou administratifs, le développement de la communication numérique et des outils d'aide à la traduction, et, plus récemment, l'arrivée massive de la traduction automatique, ont profondément bouleversé la notion de qualité. Le caractère fonctionnel de la traduction pragmatique a tout d'abord recentré la notion de qualité sur la prise en compte de l'usage final du document traduit. Les outils de traduction assistée par ordinateur (TAO) et le partage des mémoires de traduction ont ensuite dilué la responsabilité finale du traducteur dans la qualité des traductions produites, désormais déterminée, en amont, par la qualité des bases terminologiques et des segments pré-traduits et, en aval, par le contrôle de qualité et/ou la révision assurés par des relecteurs et réviseurs humains ou par des outils de contrôle qualité. La traduction automatique, enfin, fait de l'opérateur humain, soit un

simple « valideur » de solutions linguistiques produites par la mise en œuvre d'algorithmes mathématiques, soit, dans le meilleur des cas, un « post-éditeur », révisant des traductions réalisées par un moteur de traduction, afin d'obtenir un niveau de qualité donné. Elle induit, de ce fait, deux effets en apparence contradictoires : en rétablissant, d'une part, la notion de norme, à travers la traduction « de référence » qui permet de mesurer quantitativement les écarts de la traduction produite par le moteur de traduction ; en introduisant, d'autre part, la notion de relativité de la qualité, qui se mesure désormais en fonction de son degré d'acceptabilité par l'utilisateur final. Ainsi s'opère un glissement de la notion de qualité absolue, définie par rapport à une norme « extérieure » à la traduction, vers la notion de seuil d'acceptabilité, défini par rapport à une utilisation ou à une attente particulières (Bowker & Ehgoetz, 2007).

- 2 La croissance rapide du marché de la traduction et sa mondialisation à partir du milieu des années 1990 ont fait de la qualité en traduction un enjeu professionnel et commercial majeur. L'exigence de qualité s'est tout d'abord appliquée au processus mis en œuvre, à travers les différentes normes d'assurance de qualité développées depuis vingt ans (DIN 2345, ISO 9001, EN 15038, ISO 17000), qui cherchent à définir les conditions optimales de réalisation d'une traduction professionnelle. Mais la qualité du processus vise avant tout à garantir pour le donneur d'ordre et le destinataire final celle du produit lui-même : la traduction. C'est celle qui nous intéressera ici. Pour le traducteur, la qualité de ses traductions dans les conditions imposées par le donneur d'ordre est l'une des clés du maintien de son activité professionnelle. Pour l'entreprise de services linguistiques, c'est l'optimisation du rapport qualité/prix/délais qui détermine son accès aux marchés, en fonction de sa capacité à jouer sur chacune des trois variables selon les conditions commerciales et les exigences du donneur d'ouvrage. Pour le concepteur et développeur d'outils de traduction automatique, la qualité des traductions produites est l'élément essentiel permettant de départager les moteurs et les applications. Enfin, pour le formateur de traducteurs, la capacité à faire prendre conscience aux étudiants de la notion de variabilité de la qualité en contexte professionnel, tout en maintenant des objectifs de qualité maximale dans la formation et dans l'évaluation, est devenue une problématique majeure de la formation. Pour tous les acteurs, l'évaluation de la qualité des traductions est donc devenue la question cruciale, mais non résolue, évoquée dans la citation introductive.
- 3 Dans cet article, nous commencerons par examiner les réponses apportées à cette problématique par différents acteurs du monde de la traduction : formateurs de traducteurs, professionnels de la traduction et de la localisation et développeurs de solutions automatiques ou semi-automatiques. Nous proposerons ensuite une méthodologie et une grille d'évaluation multifactorielle qui permette de tenir compte des atouts et lacunes des grilles existantes. Nous illustrerons son intérêt par un exemple détaillé montrant son utilisation en contexte pour évaluer un étudiant en traduction. Enfin, nous envisagerons différentes perspectives d'utilisation de la grille dans un contexte professionnel comme dans le cadre de la formation des traducteurs.

1. Évaluation de la qualité en traduction spécialisée aujourd'hui

- 4 Selon Louise Brunette et Paul Horguelin (1998, p. 6), l'évaluation se définit généralement comme l'« examen attentif d'une traduction achevée afin d'obtenir une notation chiffrée ».
- 5 Cette définition s'applique au contexte professionnel comme au contexte pédagogique. Même si, dans le monde professionnel, l'évaluation peut parfois constituer un contrôle intermédiaire de la qualité plutôt que l'examen de la traduction « achevée », elle peut y avoir trois fonctions :
 - une fonction de contrôle de la qualité de traductions en cours de réalisation (par échantillonnage) ;
 - une fonction de sélection des prestataires de services de traduction (recrutement de traducteurs sous-traitants ou salariés) ;
 - une fonction d'aide à la décision pour déterminer si une traduction doit être acceptée et rémunérée, en fonction de son niveau de satisfaction à des critères de qualité prédéfinis.

1.1. Dans le milieu universitaire

- 6 À l'université, l'évaluation est essentiellement utilisée dans un but de formation et de mesure du niveau et des progrès de l'étudiant en traduction. Dans les travaux de recherche sur l'évaluation, les modèles s'inspirent plus ou moins de la pratique professionnelle et sont plus ou moins quantitatifs/qualitatifs ou empiriques.

1.1.1. L'évaluation dans le cadre de l'enseignement

- 7 Dans les formations de traducteurs, comme dans d'autres formations, la difficulté de parvenir à un modèle idoine d'évaluation tient à la dualité entre le besoin de sanctionner un certain niveau de compétences par des points (ou des appréciations) et celui d'assurer la progression des étudiants en justifiant l'évaluation de façon objective et circonstanciée.
- 8 En traduction, en particulier, les modèles d'évaluation « sommative » basés sur une soustraction de points-fautes et sur des catégories d'erreurs issues des cours de thème/version généraux ou littéraires ont longtemps prévalu et prévalent encore dans certaines formations.
- 9 À l'inverse, la nécessité d'offrir une évaluation « formative » aux futurs traducteurs a poussé un certain nombre d'universitaires à appliquer des principes plus holistiques ou minimalistes, basés sur une opposition entre transfert du sens et qualité linguistique, ou encore sur des théories sociolinguistiques ou argumentatives (Pym, 2003 ; House, 1997 ; Williams, 2001).
- 10 Face à la diversité des pratiques dans l'enseignement de la traduction, quelques chercheurs ont tenté de recentrer l'évaluation sur des critères plus structurés ou scientifiques : Gouadec, par exemple, a introduit la prise en compte de la cause et de la conséquence de l'erreur (ou « distortion »), qui a abouti à un système d'évaluation basé sur 675 types d'erreurs (Gouadec, 1981). Martínez et Hurtado, quant à elles, ont effectué

un état des lieux de la question de l'évaluation et de la compétence en traduction (Martínez 1997 ; Martínez & Hurtado, 2001 ; PACTE, 2011).

- 11 Ces travaux majeurs n'ont, cependant, pas permis de parvenir à un modèle d'évaluation universitaire suffisamment fidèle aux contraintes professionnelles de la traduction, tout en restant simple d'utilisation et adapté à la mesure des progrès et à la définition de profils d'étudiants traducteurs.

1.1.2. La recherche sur l'évaluation des traductions

- 12 Les travaux de recherche sur l'évaluation des traductions professionnelles et universitaires, quant à eux, ont accompagné (ou précédé) l'évolution de l'évaluation dans l'enseignement, délaissant progressivement les modèles d'évaluation littéraire ou biblique (basés sur la fidélité à l'auteur) pour se concentrer sur des approches plus en phase avec les contraintes professionnelles.
- 13 En témoignent ainsi la quantité de travaux reprenant le modèle SICAL d'évaluation des traductions professionnelles au Canada (Larose, 1998 ; Williams, 2001 ; Martínez, 2001 ; Dragovic-Drouet, 2003) ou prolongeant la théorie du *Skopos* ou théorie fonctionnaliste de Vermeer (Nord, 1997 ; Kussmaul, 1995).
- 14 D'autres recherches ont mis l'accent sur la compréhension du processus cognitif inhérent à la traduction (Jakobsen, 2011 ; Carl, 2011), les plus récentes et prometteuses rapprochant l'évaluation du produit de la traduction de son processus.
- 15 Cependant, là encore, les modèles et les orientations sont multiples, et aucun système ne parvient, pour l'instant, à rapprocher de façon satisfaisante les exigences de la traduction professionnelle et de la formation des futurs traducteurs.

1.2. Dans le milieu professionnel

1.2.1. L'évaluation en traduction et localisation

- 16 Les modèles utilisés couramment en traduction, et surtout en localisation par les professionnels (agences et fournisseurs de services linguistiques), sont en effet essentiellement quantitatifs et reposent sur le relevé des erreurs dans une traduction donnée. Quelles que soient la qualité globale d'une traduction et la justesse de certaines phrases ou segments, ce sont en effet les erreurs, leur nombre, leur nature et leur gravité qui vont déterminer l'acceptation par le donneur d'ordre de la traduction livrée.
- 17 Les principaux fournisseurs de services linguistiques mondiaux s'appuient sur des grilles d'évaluation basées sur des approches et des critères assez proches (O'Brien, 2011). Les erreurs sont relevées par un évaluateur qui les classe ensuite dans diverses catégories, selon leur nature et leur gravité. Certaines de ces grilles ont été spécialement développées par des acteurs mondiaux du secteur, en tant qu'outils d'évaluation propriétaires. La grille d'un des leaders du marché, par exemple, propose les catégories suivantes : Cohérence, Langue, Terminologie, Style, Format, Localisation, Conformité. Intervient ensuite le niveau de gravité de chaque erreur : critique, important, mineur, ou modification préférentielle de l'évaluateur. Sur le même modèle, la grille d'évaluation d'un autre acteur majeur du secteur, quant à elle, propose les catégories d'erreur suivantes : Cohérence, Format, Style, Localisation, Langue, Terminologie qui se déclinent ensuite, suivant leur niveau de gravité, sur trois niveaux : majeur, mineur ou préférentiel.

- 18 En fonction des erreurs relevées, ces grilles calculent automatiquement un score pondéré qui, rapporté au nombre de mots du texte source ou de l'échantillon étudié, détermine un indice de qualité (TQI ou *Translation Quality Index*), généralement sous forme de pourcentage, qui permet de définir si la qualité de la traduction est considérée comme acceptable ou non (seuil *pass/fail*). Une section dédiée aux commentaires permet d'apporter des précisions sur la qualité générale de la traduction, ses points forts et ses axes d'amélioration. Le deuxième exemple de grille mentionné propose, en outre, un système de bonus/malus qui permet d'augmenter ou de minorer de quelques points le score obtenu si le sentiment général est meilleur ou moins bon que le résultat mathématique des points attribués. Ce système de bonus/malus est normalement réservé aux cas extrêmes et proches de la limite d'acceptation ou de refus de la traduction.
- 19 D'autres grilles ont été développées par des associations professionnelles. La plus connue est sans doute la grille LISA (Localisation Industry Standards Association QE model), développée par les nouvelles entreprises de localisation de logiciels dans les années 1990 et diffusée (version 3.1) jusqu'à la disparition de l'association en mars 2011. La grille LISA prend en compte des catégories d'erreurs spécifiquement adaptées à son secteur d'origine, distinguant notamment les erreurs en fonction du type de support (Documentation, Aides en lignes, Logiciels) dans lequel elles se trouvent. Là encore les niveaux de gravité des erreurs sont au nombre de trois : mineur, majeur et critique, avec une pondération en fonction du niveau. Autre système d'évaluation largement cité et diffusé depuis une dizaine d'années ; la grille « SAE J2450 Translation Quality Metric », développée par la *Society of Automotive Engineers*, pour l'évaluation des traductions techniques dans le domaine de l'automobile. Fondée sur sept catégories d'erreurs (Terminologie, Syntaxe, Omission, Structure lexicale, Orthographe, Ponctuation, Divers) et deux niveaux de gravité (majeur et mineur), la grille SAE J2450 a l'avantage de la simplicité, mais l'inconvénient d'être essentiellement adaptée à un seul type de document et de traduction.
- 20 Toutes ces grilles ont en commun d'être bi-dimensionnelles (type d'erreur/gravité) et de se traduire par une acceptation ou un refus de la traduction (*pass/fail*), une forme adaptée à un usage professionnel. Dans ce cadre, elles servent, avant tout, au recrutement de sous-traitants ou de collaborateurs, à l'établissement de grilles de rémunération des sous-traitants et dans le processus de contrôle de la qualité, en cours de traduction ou avant livraison. De par leur nature et leur fonction, elles ne présentent que peu d'intérêt pour la formation, que ce soit pour l'autoformation des traducteurs en exercice ou pour la formation de futurs traducteurs en cours d'étude.
- 21 L'analyse exhaustive des grilles professionnelles en usage dans les « industries de la langue », menée par O'Brien dans le cadre du « TAUS benchmarking exercise » (2011) confirme les limites de ces grilles. O'Brien note en particulier que peu de grilles prennent en compte, de façon systématique et surtout explicite, l'effet des erreurs sur la fonctionnalité de la traduction et son utilisation par le destinataire final (même si le modèle LISA v.3.1 suggère de prendre en compte l'impact de la qualité de la documentation sur l'utilisateur). O'Brien conclut dans « JosTrans » en 2012 :

*The TAUS QE benchmarking exercise demonstrated that the preferred method for evaluating translation as a product in the translation industry is the error typology, with associated penalties and severity levels. This model, while appropriate in some contexts, cannot cater well for emerging content types, various communication channels and new needs. A more dynamic approach to QE seems to be needed by at least some members of the translation production sector.*¹ (TRADUIRE O'Brien, 2012, p. 20)

1.2.2. Évaluation et traduction automatique

- 22 La problématique de l'évaluation de la qualité de la traduction tend à se complexifier encore dès lors qu'elle s'applique à des systèmes de traduction entièrement automatiques.
- 23 Paradoxalement, les critères d'évaluation utilisés dans ce cadre sont généralement très simplifiés comparés à ceux qui s'appliquent habituellement à la traduction humaine et se fondent essentiellement sur le paradigme suivant : la qualité d'une traduction obtenue par un système de traduction automatique s'évalue par comparaison avec une traduction de référence produite par un humain et cette comparaison peut s'effectuer automatiquement. La plupart des méthodes d'évaluation s'appuie alors sur deux composantes principales : une composante d'appariement permettant d'assurer que la traduction générée automatiquement comporte les bons « mots » ; une composante d'ordre permettant de mesurer à quel point ces « mots » sont à la même place que dans la traduction de référence. Parmi les nombreuses mesures qui ont déjà été proposées et qui s'appuient sur ce paradigme, nous pouvons notamment citer le score BLEU (Papineni *et al.*, 2002).
- 24 Ce processus d'évaluation a déjà fait l'objet de nombreuses critiques dans la littérature (Wisniewski *et al.*, 2013). Certaines sont liées à la méthodologie d'évaluation proprement dite. Pour constituer les traductions de référence, les juges humains s'accordent sur une notation qui s'appuie uniquement sur deux critères : l'une porte sur la notion d'adéquation (qui mesure à quel point le sens est conservé dans la traduction) ; la seconde, sur la fluidité (qui mesure la correction de la phrase du point de vue de la langue cible, à savoir la syntaxe ou la lisibilité). Ces deux notes sont des variables ordinales qui correspondent à des jugements qualitatifs assez élémentaires (très bonne traduction, bonne traduction, mauvaise traduction, etc.). Bien que ces critères soient pertinents, ils demeurent nettement insuffisants pour évaluer efficacement la qualité d'une traduction.
- 25 Une autre critique souvent formulée à l'encontre des métriques d'évaluation automatiques porte sur les critères utilisés pour comparer la traduction de référence et la traduction automatique. Ces critères ne permettent pas en effet de rendre compte de la variabilité de la traduction (les nombreuses manières d'exprimer un sens identique dans la langue cible), des déplacements ou des phénomènes de transposition (distinguer les déplacements de mots ou de groupes de mots qui modifient le sens de la phrase de ceux qui laissent le sens inchangé) ou du fait que tous les mots n'ont pas la même importance dans la compréhension de la phrase et que cette importance peut changer en fonction du contexte applicatif.
- 26 Pour contourner ces limites, de nouvelles métriques d'évaluation automatique telles que Meteor, TERp ou hTer (Snover *et al.*, 2009) ont été proposées. Elles visent notamment à intégrer des ressources linguistiques externes pour tenter de mieux prendre en compte le phénomène de paraphrase.
- 27 D'une manière plus générale, les processus d'évaluation mis en place pour estimer la pertinence des sorties générées par les systèmes de TA restent assez limités et constituent seulement une première étape destinée à choisir le système le plus performant. À ce premier processus d'évaluation vient se greffer le travail (parfois important) de post-édition du traducteur humain qui doit également à son tour être évalué, en s'appuyant

sur des critères qui peuvent être encore différents de ceux utilisés pour évaluer la qualité d'une traduction humaine (Robert, 2010).

- 28 Il est évident, au vu de ce qui précède, que les méthodes actuelles d'évaluation de la qualité des traductions automatiques, même post-éditées, ne répondent en rien aux besoins d'une analyse « dynamique » telle que la préconise O'Brien, et ne permettent pas de concilier une évaluation à la fois professionnelle et didactique telle que nous la préconisons.

2. Proposition de grille adaptable à différents contextes et critères d'acceptabilité

- 29 Notre démarche se situe à la convergence des pratiques professionnelles et de la didactique de la traduction. Elle se situe dans la lignée d'une conception fonctionnaliste de la traduction, vue comme un produit destiné à un public et à un usage donnés, mais prenant en compte tous les facteurs de variabilité potentiels définis par Gouadec (2001, 2007) en termes de coût, d'efficacité, d'accessibilité, de registre et de qualité finale.
- 30 Nous adopterons ici les définitions suivantes :
- « Erreur » : tout écart de traduction par rapport à une norme de qualité absolue ou relative déterminée par l'évaluateur de la traduction pour un segment cible donné. Une « erreur » peut être d'ordre linguistique, sémantique ou formel ;
 - « Effet » : l'impact de l'erreur sur la qualité de la traduction, en fonction de la nature et de l'objectif de cette dernière ;
 - « Gravité » : l'importance de l'effet de l'erreur.

2.1. Genèse

- 31 Notre approche de l'évaluation de la qualité est née d'une expérimentation visant à comparer la productivité de plusieurs types d'outils d'aide à la traduction (outil de TAO, traduction automatique et système de reconnaissance vocale), appliqués à la traduction éditoriale « pragmatique » (traduction d'ouvrages informatifs, didactiques ou instructifs) par des étudiants du master « Métiers de la traduction-localisation et de la communication multilingue et multimédia » du CFTR Rennes 2². Il s'agissait d'étudier la corrélation entre les temps de traduction obtenus avec les différents outils et la qualité des traductions produites, afin de déterminer l'outil le mieux à même de produire un niveau de qualité donné. L'évaluation qualitative reposait sur une grille standard d'évaluation de type professionnel, telle que décrite plus haut, fondée sur le relevé des types d'erreur (omission, erreur de sens, style, terminologie, orthographe, grammaire, ponctuation) et de leur niveau de gravité (mineur, majeur, critique) aboutissant au calcul d'un indice de qualité (TQI) à partir du rapport nombre de mots évalués/nombre d'erreurs.
- 32 Cette première expérimentation n'ayant pas fourni de résultats suffisamment probants, notamment en raison de la disparité de niveau des traducteurs testés, nous avons été rapidement amenés à réfléchir, d'une part sur les objectifs de cette étude et, d'autre part, sur l'élaboration d'une grille d'évaluation plus fine, permettant de mieux prendre en compte la variabilité inhérente aux situations de traduction.

- 33 Partant du constat que les traducteurs que nous pouvions solliciter pour l'expérimentation étaient des traducteurs en formation, nous avons donc décidé d'inclure parmi les objets de notre étude l'effet des outils et des méthodes sur les performances des apprentis-traducteurs, tout en maintenant l'objectif d'une évaluation qualitative fondée sur des critères professionnels.
- 34 Cette réflexion a mené à l'élaboration d'une grille d'évaluation multicritères qui devait pouvoir répondre à plusieurs objectifs :
- s'appliquer aussi bien à des traducteurs professionnels qu'à des traducteurs en formation ;
 - s'appliquer à tout type de traduction impliquant une intervention humaine, quelle que soit sa nature (traduction initiale, relecture/révision ou post-édition d'une traduction machine) ;
 - permettre une pondération différente des défauts de qualité en fonction du type de texte source et de l'objectif du document traduit ;
 - limiter, autant que possible, les biais subjectifs ;
 - être facile d'utilisation et permettre le calcul automatique d'un score.
- 35 Comme aucune des grilles d'évaluation existantes ne permettait de répondre exactement à ces critères (O'Brien, 2012), nous avons donc réfléchi à l'élaboration d'une grille dédiée, pouvant croiser plusieurs critères de qualité et permettant une analyse plus fine et plus dynamique que les grilles standard.

2.2. Principes

- 36 Ne souhaitant pas « réinventer la roue », nous avons pris pour point de départ les grilles d'évaluation professionnelles en vigueur, fondées sur une typologie des erreurs de traduction.
- 37 L'originalité et le potentiel de la grille d'évaluation que nous proposons tiennent dans le croisement de ce critère de qualité avec deux autres dimensions : d'une part, l'effet de l'erreur sur la qualité de la traduction, en fonction de la nature de celle-ci, de son objectif, de ses destinataires et de son usage ; d'autre part, le degré de gravité de l'effet induit par l'erreur (et non de l'erreur en elle-même), sur une échelle de 0 à 3. C'est le nombre de points attribués à la gravité de chaque effet qui est ensuite totalisé pour produire un score global sur l'échantillon évalué. Plus ce score est élevé, moins la qualité est conforme au critère adopté.
- 38 Du fait du caractère tridimensionnel de la grille, chaque type d'erreur peut donc avoir un effet multiple, permettant de tenir compte à la fois de la nature de la traduction évaluée et de l'objectif de l'évaluation.

2.2.1. Typologie des erreurs

- 39 Nous distinguons dans la grille neuf types d'erreurs : sept d'entre eux reposent sur des catégories traditionnelles (Sens, Grammaire/syntaxe, Orthographe/typographie, Terminologie, Phraséologie, Style, Omission/ajout), auxquels s'ajoutent deux catégories issues de l'évaluation professionnelle : les erreurs de localisation (définies comme l'absence d'adaptation à un public cible ou à une culture donnée) et les erreurs de PAO (défauts de mise en page et de formatage).

- 40 Nous reproduisons ci-dessous l'ensemble des explications concernant la typologie des erreurs disponibles dans la feuille 3 de la grille et auxquelles l'évaluateur peut se référer à tout moment en cas de doute.

Sens	Omission/addition	Terminologie	Phraséologie
Ambiguïté sémantique	Omission d'un élément significatif du texte source modifiant le sens du texte cible	Variante inappropriée (variété de langue/usage professionnel/usage « maison »)	Variante inappropriée (variété de langue/usage professionnel/usage « maison »)
Perte partielle de sens	Ajout injustifié d'informations ayant un impact mineur sur le texte cible	Terme inapproprié appartenant à un autre domaine	Phraséologie inappropriée appartenant à un autre domaine
Perte totale de sens	Ajout injustifié d'informations ayant un impact majeur sur le texte cible	Incohérence terminologique (dans le texte cible/par rapport aux documents de référence)	Incohérence phraséologique (dans le texte cible/par rapport aux documents de référence)
Incohérence manifeste du texte source non signalée ou rectifiée dans la traduction			

Grammaire/syntaxe	Orthographe/typographie	Style	Localisation	PAO
Erreurs morpho-syntaxiques	Orthographe non conforme	Calque sur le texte source	Non adaptation à la culture cible	Mise en page
Ordre des mots	Fautes de frappe	Longueur des phrases	Non adaptation au public cible	Formatage
Structure de la phrase	Ponctuation et typographie non conformes	Maladresses, manque de fluidité	Pas de localisation des données chiffrées	Graphiques
		Registre de langue inapproprié (formel/informel)		Balises
		Variété de langue inappropriée (orthographe ou lexicale)		Références croisées

2.2.2. Typologie des effets

- 41 Cette deuxième dimension repose sur quatre critères. Deux de ces derniers sont communs aux grilles d'évaluation de la traduction automatique, à savoir la « précision » (*accuracy*) et la « lisibilité » (*fluency*), même si notre définition de ces critères est plus restrictive que celle de Koehn (2007). Deux autres sont empruntés à l'univers de la traduction professionnelle « humaine », à savoir la « fonctionnalité », définie comme la capacité de la traduction à remplir la fonction qui lui est assignée (informer, donner des instructions, mettre en garde, etc.) et la « conformité », définie comme le respect d'une norme explicite ou implicite (feuille de style, cahier des charges, terminologie imposée, convention linguistique et/ou culturelle, etc.).
- 42 La définition des effets est résumée ci-dessous :

Précision	Fonctionnalité	Lisibilité	Conformité
Erreur empêchant la traduction de transmettre l'information présente dans le document source	Erreur empêchant la traduction de remplir la fonction qui lui est assignée	Erreur nuisant à la fluidité du document cible et à sa compréhension par l'utilisateur final	Erreur entraînant la non conformité aux normes, conventions ou recommandations reconnues ou imposées par le donneur d'ordre

2.2.3. Niveau de gravité

- 43 La troisième dimension de la grille est celle du niveau de gravité de l'effet induit par l'erreur constatée. Elle permet de distinguer quatre niveaux de gravité affectant la qualité du document cible : 0 = pas d'effet/effet non comptabilisé, 1 = mineur, 2 = majeur, 3 = erreur critique. Ces niveaux de gravité sont rattachés aux quatre types d'effets que nous venons d'explicitier et non aux neuf types d'erreurs, comme c'est le cas dans la plupart des modèles d'évaluation professionnels. C'est donc bien la conséquence de l'erreur qui est ici sanctionnée et non sa cause.

2.3. Mise au point et ajustement

- 44 En partant des critères définis ci-dessus, la grille d'évaluation a été testée sur les traductions effectuées dans le cadre d'une expérimentation contrôlée de traduction réalisée en février et mars 2013. Les traductions d'un même document anglais de 500 mots (notice d'installation d'un système de chauffage solaire pour piscines), effectuées avec trois outils différents par un groupe d'étudiants de master 2, ont été évaluées à l'aide de la grille décrite ci-dessus.
- 45 En vue de limiter la part de subjectivité de l'évaluation humaine, nous avons affecté quatre évaluateurs pour chaque traduction dans la première phase d'utilisation de la grille, pour n'en garder que deux dans des phases ultérieures, les évaluateurs étant alors plus aguerris à l'utilisation de l'outil. Cette confrontation de nos évaluations présentait un double avantage : elle nous a permis d'une part — tout particulièrement pour le premier lot qui servait ainsi de phase de pré-test de la grille — d'affiner la grille et les définitions de chaque type d'erreur et des effets induits. Nous avons notamment ajouté une colonne « correction » afin, principalement, de faciliter le travail de coordination entre les différents évaluateurs. Le besoin de cette colonne s'est rapidement fait ressentir pour indiquer clairement quelle erreur était concernée dans le passage cité et pour vérifier rapidement si son catalogue était correct. Dans cette optique, nous avons également décidé de fournir un contexte suffisant dans le document source pour cerner l'erreur et de souligner en rouge le ou les mot(s) qui posent problème dans la traduction. Rappelons que, au-delà de la richesse et de la précision visée par cette grille, l'objectif est également de permettre une évaluation rapide et simple des traductions.
- 46 D'autre part, la concertation nous a conduits à rechercher une évaluation plus objective (grâce à l'examen des erreurs récurrentes et de la façon dont chaque évaluateur a classé l'erreur, son effet et sa gravité), afin de réduire les différences d'un évaluateur à l'autre. Cette étape était essentielle afin d'obtenir des données comparables pour la phase d'analyse ultérieure des résultats.
- 47 La comparaison des scores attribués à chaque échantillon lors du traitement des premiers lots a permis de s'interroger sur les disparités entre évaluateurs. Il s'est avéré rapidement

que l'une des sources de divergences était à chercher dans la répartition différente des points entre les différents critères et des disparités dans l'évaluation de la gravité. Il a donc été décidé :

- de limiter à 5 le nombre de points de pénalité pour une erreur donnée ;
 - d'éviter la dispersion des types d'effet entre les quatre critères (attribution d'une gravité mineure sur chacun des quatre effets, par exemple). Pour y remédier, nous avons décidé, après concertation, de limiter le nombre d'effets à 2 (sur 4) en déterminant obligatoirement un effet dominant, auquel il est possible d'adjoindre un seul effet secondaire. À titre d'exemple, une erreur peut entraîner une pénalité de deux points imputés à l'effet « fonctionnalité » et d'un point à l'effet « précision », mais ne peut plus être pénalisée au niveau des deux autres types d'effets (lisibilité et conformité).
- 48 Au terme de cette période de test, nous avons estimé que la grille était opérationnelle et pouvait être testée à plus grande échelle.

2.4. Exemple d'évaluation à l'aide de la grille

- 49 Afin d'offrir une représentation plus concrète de la grille TRASILT, nous reproduisons ci-dessous un exemple d'évaluation de la traduction d'un étudiant réalisée à l'aide de la grille.

Résultats de l'évaluation			
Informations (document)		Informations (traducteur)	
Nom du document cible	Solarhot Installation/Owner Manual	Traducteur	[...]
Numéro de lot	Trad5_TA-GT	Code opérateur	BP
Numéro de page	2	Équipe	
Nombre de mots sources	527	Niveau	
Nombre de mots cibles	571		
Langue source	Anglais		
Langue cible	Français		

Texte source	Traduction fautive	Type d'erreur	Niveau de gravité (0 = pas d'effet/effet non comptabilisé, 1 = mineur, 2 = majeur, 3 = critique)				Correction
			Précision	Fonctionnalité	Lisibilité	Conformité	
There are air bubbles	Il ya des bulles d'air	Orthographe/typographie	0	0	0	1	Il y a des bulles d'air
There are air bubbles in the pool only when the solar is operating	Il ya des bulles d'air dans la piscine lorsque l'énergie solaire est en marche	Omission/addition	1	0	0	0	seulement lorsque l'installation solaire
when the solar is operating	lorsque l' énergie solaire est en marche	Sens	3	2	0	0	système de chauffage solaire
due to the pump working harder to move	La pompe fonctionne alors de façon plus importante pour déplacer l'eau dans le système solaire	Style	0	0	1	0	la pompe fonctionne alors en surrégime
to move the water through the solar system	pour déplacer l'eau dans le système solaire	Sens	0	2	0	3	système de chauffage solaire
as it flows by the valve	qui s'écoule par le robinet	Sens	3	0	0	0	au niveau de la vanne
as it flows by the valve	qui s'écoule par le robinet	Terminologie	0	0	0	2	au niveau de la vanne
Check the "O" ring	Vérifiez l' anneau en "O"	Terminologie	1	2	0	0	Vérifiez le joint torique
Clean, lubricate or replace as needed	Nettoyez, graissez ou remplacez- le	Grammaire/syntaxe	0	0	0	1	nettoyez, graissez ou remplacez le joint

3. Perspectives d'utilisation et d'amélioration de la grille

- 50 Comme le démontre l'exemple présenté, la grille permet donc une analyse précise de l'erreur et de son poids sur la qualité du document produit, tout en offrant une souplesse d'adaptation de la pondération grâce à une décorrélation entre la classification de l'erreur et son effet sur la qualité.

3.1. Dans la formation des traducteurs

- 51 Pour l'enseignement de la traduction dans le cadre de formations professionnalisantes, la grille permet, en outre, d'exercer les étudiants à tout type de document source ou de prestation de traduction, grâce à l'exhaustivité des types d'erreurs proposés (catégories traditionnelles, mais aussi localisation ou PAO) et à la mise en valeur de l'effet des erreurs sur tel ou tel critère de qualité lié à la spécialité, au type de document ou à la commande de traduction.
- 52 On peut ainsi imaginer utiliser la grille pour privilégier un aspect particulier de la qualité lié au type de document, en spécifiant dans un exercice l'intolérance aux erreurs de phraséologie dans le cadre de la traduction d'un contrat, par exemple.
- 53 Elle permet également de dépasser la question des outils d'aide à la traduction éventuellement employés pour l'exercice, pour se concentrer sur le résultat atteint par l'étudiant. Ceci n'empêche cependant pas d'observer des tendances sur les types d'erreurs récurrents induits par telle ou telle méthode, et de faire travailler les étudiants à partir d'un constat chiffré (le nombre d'erreurs de style induites par la post-édition d'une traduction automatique, par exemple).

- 54 La grille permet également de discriminer plus précisément les erreurs commises chez un même étudiant et d'en déterminer les récurrences, si elle est utilisée de façon systématique par le formateur. Ceci peut permettre d'observer la progression de l'étudiant de façon précise et détaillée. La grille peut d'ailleurs être complétée par une zone de commentaires sur la qualité globale de la traduction produite par l'étudiant, afin de compléter la fonction formative de l'évaluation.
- 55 La grille peut encore évoluer en identifiant en amont les caractéristiques qualitatives du document source sur lequel travaillent les étudiants (difficultés du document, problèmes de rédaction, type de discours, etc.). On peut enfin imaginer une colonne supplémentaire dans la grille, qui permettrait d'élucider la cause probable de chaque erreur. L'hypothèse d'explication pourrait être formulée par l'enseignant ou encore par l'étudiant lui-même après évaluation de l'enseignant.
- 56 Cette grille est logiquement le support idéal pour la pratique universitaire du contrôle de qualité en suivant des critères professionnels. Elle permet également de servir à l'enseignement de la révision en se concentrant davantage sur l'amélioration qualitative plutôt que sur le comptage des points.

3.2. Dans le milieu professionnel

- 57 Quels seraient les avantages d'une telle grille dans l'évaluation des traductions professionnelles ? L'avantage principal que nous y voyons est sa souplesse, permettant, selon la nature des traductions réalisées et l'objectif de l'évaluation, de varier la pondération accordée à tel ou tel type d'erreur ou tel ou tel critère fonctionnel dans le calcul du score final.
- 58 Ainsi peut-on imaginer que l'évaluation de la traduction d'un rapport financier ou d'un article de recherche en sciences expérimentales accordera un poids plus fort au facteur « précision ». La traduction d'une notice technique ou d'un manuel utilisateur nécessitera de privilégier plus particulièrement le facteur « fonctionnalité ». Un document promotionnel, quant à lui, sera évalué prioritairement à l'aune de sa lisibilité. Des tests de traduction destinés à sélectionner des collaborateurs ou des sous-traitants pourront mettre l'accent successivement sur plusieurs critères : la précision pour mesurer la maîtrise de la langue source, la fonctionnalité pour évaluer la capacité du traducteur à appréhender le fonctionnement d'un système et à le décrire clairement pour l'utilisateur, la lisibilité pour mesurer la fluidité du style du traducteur et sa maîtrise de la langue cible, et la conformité pour vérifier sa capacité à respecter un guide de style ou des ressources terminologiques ou phraséologiques fournies au préalable.
- 59 La grille proposée représente, selon nous, un outil utilisable en l'état par les professionnels de la traduction et du contrôle de la qualité, moyennant un léger entraînement de l'évaluateur aux critères de la grille, mais son utilisation peut être accélérée grâce à quelques améliorations d'ordre technique.

3.3 Amélioration des fonctionnalités de la grille

- 60 Bien que le processus d'évaluation de la qualité d'une traduction soit difficilement automatisable ou à un niveau de qualité très réduit (l'exemple de l'évaluation automatique des systèmes de traduction automatique en témoigne), nous pouvons

néanmoins proposer quelques améliorations techniques destinées à faciliter la tâche de l'évaluateur.

3.3.1. Recopie automatique des phrases erronées et sélection du type d'erreurs

61 Dans son travail, l'évaluateur est confronté à de nombreuses manipulations de documents aux multiples formats (pdf, document texte, tableur, etc.). Il doit en effet lire les documents source et cible, repérer les erreurs, les recopier dans la grille d'évaluation (généralement au format tableur), y insérer des commentaires et des suggestions de corrections, puis sélectionner les types d'erreurs, leur gravité, etc. Pour réduire le nombre de ces opérations, nous proposons d'automatiser une partie du processus en développant une macro-commande (programme informatique) qui accompagne l'évaluateur de la manière suivante :

- pour chaque erreur dans le document cible, l'évaluateur surligne la phrase concernée. La couleur du surlignage sera choisie parmi une palette de couleurs prédéfinie en fonction du type d'erreurs. Un surlignage en jaune indique, par exemple, une erreur de sens (*meaning*), un surlignage en rouge une erreur de style (*style*), etc. ;
- la macro-commande développée permet ensuite de recopier automatiquement et une à une dans la grille d'évaluation (colonne « Traduction fautive ») les phrases surlignées par l'évaluateur dans le document cible et de sélectionner automatiquement le type d'erreurs en fonction de la couleur du surlignage. L'évaluateur termine son travail en ajoutant ses commentaires et corrections, puis en sélectionnant les niveaux de gravité des erreurs relevées.

3.3.2. Calcul automatique du score global

62 Après la phase d'évaluation de la qualité de la traduction, nous obtenons une liste d'erreurs, classées par type, et associées à un niveau de gravité lié à l'effet de l'erreur sur la qualité (voir exemple décrit plus haut).

63 L'ultime étape de l'évaluation consiste à obtenir un score global permettant de représenter la qualité de la traduction dans son ensemble. Pour calculer ce score, nous proposons, à partir du relevé des erreurs et des scores attribués à chaque erreur individuelle dans la grille d'évaluation, de remplir automatiquement le tableau synthétique suivant :

Catégorie	Nombre d'erreurs	Précision	Fonctionnalité	Lisibilité	Conformité	Total
Sens	9	16	15	1	3	35
Omission/addition	2	1	3	0	2	6
Terminologie	3	1	2	0	2	5
Phraséologie	0	0	0	0	0	0
Grammaire/syntaxe	3	0	0	0	4	4
Orthographe/typographie	1	0	0	0	1	1
Style	3	0	1	4	0	5
Localisation	0	0	0	0	0	0
PAO	0	0	0	0	0	0
Total	21	18	21	5	12	56

- 64 Le score global est donc obtenu en additionnant l'ensemble des points de pénalités (niveau de gravité compris entre 0 et 3) attribués à chaque effet pour chaque type d'erreur. Il servira, notamment, de base pour établir un seuil d'acceptabilité de la traduction ou pour comparer le niveau entre étudiants ou traducteurs en fonction des critères établis par l'utilisateur (enseignant, chercheur, commanditaire de la traduction, etc.).

3.3.3. Adaptation automatique des pondérations en fonction des types de documents, des contraintes du donneur d'ordre et/ou des objectifs pédagogiques

- 65 Par défaut, les calculs effectués pour obtenir ce score global accordent à chaque type d'erreur et à chaque effet la même importance (le même poids). Comme indiqué précédemment, la particularité de notre grille est de pouvoir être facilement adaptable au contexte dans lequel elle sera utilisée. Dans un contexte professionnel spécifique, par exemple, la définition de la qualité d'une traduction peut varier selon le type de document ou selon des contraintes particulières spécifiées par le donneur d'ordre. Pour une utilisation pédagogique, l'enseignant peut également souhaiter, pour évaluer les travaux de ses étudiants, privilégier un type d'erreur ou un effet particulier en fonction d'un objectif pédagogique spécifique.
- 66 Le système de calcul proposé est suffisamment souple pour permettre d'attribuer des pondérations différentes selon le type d'erreur ou l'effet sur la qualité. Ainsi, il sera possible à un évaluateur de « jouer » avec les pondérations et d'indiquer par exemple que, pour un document donné, il souhaite être particulièrement attentif aux erreurs de localisation et de sens tout en choisissant d'ignorer les erreurs de style.
- 67 Toujours dans l'optique de faciliter la tâche de l'évaluateur, nous prévoyons de simplifier ce travail de choix des pondérations en proposant plusieurs schémas prédéfinis qui auront pour objectif d'attribuer automatiquement des poids aux différents types d'erreurs et d'effets en fonction du type de document à traiter, des contraintes du donneur d'ordre et/ou des intérêts pédagogiques.

Conclusion

- 68 La grille d'évaluation qualitative des traductions proposée dans le présent article cherche à dépasser les limites inhérentes aux systèmes d'évaluation monofonctionnels, destinés soit à un usage didactique, soit à un usage professionnel, soit à des fins de recherche. En combinant type d'erreur, effet de l'erreur sur l'utilisateur final et degré de gravité de l'effet produit, cette grille d'évaluation dynamique permet de prendre en compte une multiplicité de situations de traduction, dans différents contextes, qu'ils soient professionnels, pédagogiques ou scientifiques.
- 69 Par le biais des effets de l'erreur, de la criticité des effets et du jeu des pondérations possibles dans le calcul des résultats, la grille proposée peut être tout à la fois un instrument de mesure de la qualité « absolue » d'une traduction ou un outil d'évaluation du seuil d'acceptabilité par l'utilisateur final en fonction de tel ou tel critère (précision, fonctionnalité, lisibilité, conformité).
- 70 La grille ainsi mise au point est déjà utilisée dans le cadre pédagogique, au sein de cours et de projets impliquant une révision et un contrôle de la qualité à l'université Rennes 2,

auprès des étudiants du master MTL2M (Métiers de la traduction-localisation et de la communication multilingue et multimédia) du CFTR (Centre de formation des traducteurs-terminologues et rédacteurs).

- 71 En tant qu'outil de recherche, elle sert depuis 2 ans à l'évaluation scientifique de la production d'apprentis traducteurs travaillant avec différents outils d'aide à la traduction sur différents types de documents, dans différents domaines (voir notamment une première présentation de ces résultats au colloque du CRITT, en janvier 2014) (Hernandez Morin *et al.*, 2014).
- 72 Cette grille devra maintenant être testée en tant qu'instrument professionnel auprès d'entreprises de traduction et de services linguistiques, ce qui suppose un travail sur son ergonomie et sur l'automatisation d'un certain nombre de fonctions de saisie et de calcul des résultats en fonction des pondérations choisies.

BIBLIOGRAPHIE

BOWKER Lynne & EHGOETZ Melissa, « Exploring User Acceptance of Machine Translation Output: A Recipient Evaluation », dans D. Kenny et K. Ryou (dir.), *Across Boundaries: International Perspectives on Translation*, Newcastle-upon-Tyne, Cambridge Scholars Publishing, 2007, p. 209-224.

CARL Michael, « A Taxonomy of Human Translation Styles », *Translation Journal*, vol. 16, n° 2, 2011. Disponible sur <www.bokorlang.com/journal/56taxonomy.htm> (consulté le 24 février 2014).

DRAGOVIC-DROUET Mila, *Évaluation de la qualité des traductions éditoriales*, thèse de doctorat, université Paris 3, ESIT, 2003.

GOUADEC Daniel, « Paramètres de l'évaluation des traductions », *Meta*, vol. 26, n° 2, 1981, p. 99-116. Disponible sur <www.erudit.org/revue/meta/1981/v26/n2/index.html> (consulté le 21 février 2014).

—, *Stratégies de traduction entre l'anglais et le français - Essai de définition d'un modèle de formation des traducteurs*, Paris, Métiers des langues/Langues des métiers, Maison du dictionnaire, 2001, 997 p.

—, *Translation as a Profession*, Amsterdam/Philadelphie, John Benjamin Publishing Company, vol. 73, 2007, 396 p.

GREEN Spence, HEER Jeffrey & MANNING Christopher D., « The Efficacy of Human Post-Editing for Language Translation », dans *CHI 2013: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Stanford, 2013, p. 439-448. Disponible sur <<http://vis.stanford.edu/files/2013-PostEditing-CHI.pdf>> (consulté le 18 février 2014).

HERNANDEZ MORIN Katell, BARBIN Franck, MOREAU Fabienne, TOUDIC Daniel & PHUEZ-FAVRIS Gaëlle, « Translation Technology and Learner Performance: Professionally-Oriented Translation Quality Assessment with Three Translation Technologies », dans *Translation in Transition: Between Cognition, Computing and Technology* (colloque CRITT-WCRE, 30-31 janvier 2014, École de commerce de Copenhague), 2014. Disponible sur <http://bridge.cbs.dk/events/presentations/panel_6/CRITT_WCRE_Barbin.pdf> (consulté le 21 février 2014).

- HOUSE Juliane, *Translation Quality Assessment: A Model Revisited*, Tübingen, Narr, 1997, 207 p.
- HORGUELIN Paul & BRUNETTE Louise, *Pratique de la révision*, 3^e édition revue et augmentée, Brossard (Québec), Linguatex éditeur, 1998, 263 p.
- JAKOBSEN Arnt Lykke, « Tracking Translators' Keystrokes and Eye Movements with Translog », dans C. Alvstad, A. Hild et E. Tiselius (dir.), *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies*, Amsterdam/Philadelphie, John Benjamins Publishing Company, 2011, p. 37-55.
- KOEHN Philipp, HOANG Hieu, BIRCH Alexandra, CALLISON-BURCH Chris, FEDERICO Marcello, BERTOLDI Nicola, COWAN Brooke, SHEN Wade, MORAN Christine, ZENS Richard, DYER Chris, BOJAR Ondřej, CONSTANTIN Alexandra & HERBST Evan, « Moses: Open Source Toolkit for Statistical Machine Translation », dans *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, 2007, p. 177-180.
- KUSSMAUL Paul, *Training the Translator*, Amsterdam/Philadelphia, John Benjamins Publishing Company, 1995, 19 p.
- LAROSE Robert, « Méthodologie de l'évaluation des traductions », *Meta*, vol. 43, n° 2, 1998, p. 163-186. Disponible sur <www.erudit.org/revue/meta/1998/v43/n2/003410ar.html?vue=resume> (consulté le 21 février 2014).
- MARTÍNEZ MELIS Nicole, *Évaluation et traduction : cadre de recherche sur l'évaluation dans la didactique de la traduction*, thèse de doctorat, Universitat Autònoma de Barcelona, 1997.
- MARTÍNEZ MELIS Nicole & HURTADO ALBIR Amparo, « Assessment in Translation Studies: Research Needs », *Meta*, vol. 46, n° 2, 2001, p. 272-287. Disponible sur <www.erudit.org/revue/meta/2001/v46/n2/003624ar.html?vue=resume> (consulté le 18 février 2014).
- NORD Christiane, *Translating as a Purposeful Activity: Functionalist Approaches Explained*, Londres, Routledge, 1997, 154 p.
- O'BRIEN Sharon, « JosTrans », 2012. Disponible sur <www.jostrans.org/issue17/art_obrien.php> (consulté le 21 février 2014).
- PACTE, « Results of the Validation of the PACTE Translation Competence Model: Translation Problems and Translation Competence », dans *Methods and Strategies of Process Research: Integrative Approaches in Translation Studies*, Amsterdam, John Benjamins, 2011, 377 p.
- PAPINENI Kishore, ROUKOS Salim, WARD Todd & ZHU Wei-Jing, « Bleu: A Method for Automatic Evaluation of Machine Translation », dans *ACL 2002: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, 2002, p. 311-318.
- PYM Antony, « Redefining Translation Competence in an Electronic Age. In Defence of a Minimalist Approach », *Meta*, vol. 48, n° 4, 2003, p. 481-497. Disponible sur <www.erudit.org/revue/meta/2003/v48/n4/008533ar.html?vue=resume> (consulté le 25 février 2014).
- ROBERT Anne-Marie, « La post-édition : l'avenir incontournable du traducteur ? », dans *Interpreting the Future* (actes du colloque organisé par le BDU, 11-13 septembre 2009), 2010.
- SNOVER Matthew G., DORR Bonnie, MADNANI Nitin & SCHWARTZ Richard, « Ter-Plus: Paraphrase, Semantic, and Alignment Enhancements to Translation Edit Rate », *Machine Translation*, vol. 23, n° 2-3, 2009, p. 117-127.
- TAUS, « Translation Quality Evaluation is Catching up with the Times ». Disponible sur <www.taus.net/reports/translation-quality-evaluation-is-catching-up-with-the-times> (consulté le 12 juin 2014).

WILLIAMS Malcom, « Application of Argumentation Theory to Translation Quality Assessment », *Meta*, vol. 46, n° 2, 2001, p. 326-344. Disponible sur <www.erudit.org/revue/meta/2001/v46/n2/004605ar.html?vue=resume> (consulté le 25 février 2014).

—, *Translation Quality Assessment: An Argumentation-Centred Approach (Perspectives on Translation Series)*, Ottawa, University of Ottawa Press, 2004, 188 p.

WISNIEWSKI Guillaume, SINGH Anil Kumar & YVON François, « Quality Estimation for Machine Translation: Some Lessons Learned », *Machine Translation*, vol. 27, n^{os} 3-4, 2013, p. 213-238.

NOTES

1. Notre traduction : « L'audit des méthodes d'évaluation réalisé par TAUS a démontré que la méthode privilégiée pour l'évaluation de la traduction en tant que produit dans l'industrie langagière est la typologie des erreurs, associée à des pénalités et des niveaux de gravité. Ce modèle, s'il est approprié dans certains contextes, n'est pas adapté aux types de contenus émergents, à la multiplication des canaux de communication ou aux nouveaux besoins en matière d'information. Une approche plus dynamique de l'évaluation de la qualité semble nécessaire, pour une partie des acteurs du monde de la traduction au moins. »

2. Centre de formation des traducteurs-localiseurs, terminologues, rédacteurs.

RÉSUMÉS

Notre article part du constat des lacunes existantes dans l'évaluation professionnelle des traductions spécialisées comme dans l'évaluation universitaire. De multiples modèles existent : purement quantitatifs pour l'évaluation de la traduction automatique (basée sur le nombre et l'ordre des mots du texte cible) ; basés sur des totaux d'erreurs en localisation (erreurs/gravité) et dans la formation (nombre de points-fautes) ; parfois ouvertement subjectifs et empiriques (utilisable/inutilisable) dans les recherches sur la fonctionnalité des traductions. Même si ces modèles posent souvent leurs propres seuils d'acceptabilité, aucun ne permet d'évaluer une traduction en fonction de critères adaptables à différents genres, types de documents, spécialités ou situations de traduction.

Notre grille quantitative et qualitative s'appuie sur l'identification des « erreurs » de traduction telle qu'elle est pratiquée en traduction spécialisée professionnelle, en y ajoutant l'effet de l'erreur sur le document traduit. Elle est dynamique en ce qu'elle permet de corréliser ou non le type d'erreur, l'effet que l'erreur produit sur la qualité du document livré et le degré d'impact de cet effet sur l'utilisateur final de la traduction. Elle dépasse ainsi la simple étude des outils et méthodes de traduction pour se concentrer sur le produit traduit et mieux révéler les failles et les points forts de chaque traducteur en fonction de la méthode utilisée.

Utilisée et éprouvée dans le cadre de l'évaluation scientifique de traductions effectuées par des étudiants de dernière année de master appliquant différentes méthodes (traduction sans outil, avec outil de TAO, de reconnaissance vocale, en post-édition de traduction automatique), la grille peut être partiellement automatisée et s'adapter à divers contextes professionnels ou universitaires par un système de pondération des critères d'acceptabilité.

This article posits that current translation quality assessment methods applied to specialized translation, be it in a professional context or in an academic environment, remain unsatisfactory. There are currently a number of such methods, ranging from purely quantitative approaches applied to machine translation (on the basis of target text word number and order), to error and criticality criteria applied by the localization industry and in the classroom (points-based systems) and to purely subjective (satisfactory/unsatisfactory) end-user assessments in target text usability research. While all these methods often adopt their own acceptability thresholds, none really offers a way to assess a translation using a flexible set of criteria adaptable to any kind of genre, source document type, domain or situation.

Our assessment grid is designed to measure both quantitative and qualitative aspects of translation quality through commonly accepted professional error-based quality criteria, while also taking into account the effect of the error on target text functionality. The grid is designed to be dynamic by allowing error type, effect on translation quality and criticality to be assessed independently or in correlation. It aims to go beyond a simple study of tools and methods by concentrating on how the end product reveals each translator's weak or strong points according to the method used.

The grid has been tried and tested in assessing the quality of technical translations carried out in strictly controlled conditions by final year Master's degree students using different translation tools and methods (i.e. simple word processing, a translation memory system, a speech recognition tool and MT post-editing). Data entry can be partly automated and each parameter can be weighted to take into account different translation situations and acceptability thresholds.

INDEX

Mots-clés : traduction spécialisée, évaluation de la qualité, grille d'évaluation

Keywords : specialized translation, translation quality assessment, quality assessment grid

AUTEURS

DANIEL TOUDIC

Université Rennes 2, Unité de recherche LIDILE (EA3874) – TRASILT

KATELL HERNANDEZ MORIN

Université Rennes 2, Unité de recherche LIDILE (EA3874) – TRASILT

FABIENNE MOREAU

Université Rennes 2, Unité de recherche LIDILE (EA3874) – TRASILT

FRANCK BARBIN

Université Rennes 2, Unité de recherche LIDILE (EA3874) – TRASILT

GAËLLE PHUEZ

Université Rennes 2, Unité de recherche LIDILE (EA3874) – TRASILT