



# Les 12 travaux de la Traduction automatique

Journée d'études Traduction et qualité : « Biotraduction et traduction automatique ».  
Université de Lille, 2 février 2018



# Sandrine Peraldi

School of Languages, Cultures and Linguistics.  
University College Dublin

[sandrine.peraldi@ucd.ie](mailto:sandrine.peraldi@ucd.ie)

## Démocratisation galopante de la Traduction automatique (Peraldi 2016):

- ▶ la commercialisation systématique de fonctionnalités de TA dans les outils de TAO et gestionnaires de traduction,
- ▶ le nombre grandissant de colloques et publications (TRALOGY I et II, TAOCAT, Translating Europe Forum, etc.),
- ▶ le développement par la CE/DGT de EC@MT dans une volonté d'harmonisation et de facilitation du processus de traduction, à destination des différents organes de l'Union européenne.
- ▶ les progrès fulgurants de la TA (traduction neuronale).

## Adéquation des formations universitaires aux besoins du marché:

- ▶ Prise en compte de la multiplication et de la diversification des outils pour en faire une compétence professionnelle (Frérot, Karagouch 2016)
- ▶ Intégration systématique de modules dédiés à la TA, à la post-édition
- ▶ Initiation depuis plusieurs années d'une réflexion globale sur les apports technologiques en traduction (EMT, OPTIMALE, etc.)
- ▶ Mise à disposition d'un arsenal pédagogique (tutoriels, corpus, supports de cours, etc.)

## Un domaine de recherche sous-représenté :

- ▶ Très peu d'articles dans ce domaine :

*« Despite the acknowledged importance of translation technology in translation studies programmes and the current ascendancy of Statistical Machine Translation (SMT), there has been little reflection to date on how SMT can or should be integrated into the translation studies curriculum. »* (Doherty, Kenny 2014 : 295)

- ▶ État de l'art:

- ▷ Depraetere (2010, 2014)
- ▷ Korošec (2011)
- ▷ Gaspari, Almaghout, Doherty (2015)
- ▷ O'Brien (2006, 2008), Kenny (2011), Kenny & Doherty (2014), Doherty (2016)

- ▶ État de fait assez surprenant au regard des enjeux de la TA pour les futurs diplômés et d'un point de vue épistémologique (Cronin 2013, Engeström & Sannino 2010)

## Questionnements :

- ▷ Bowker, McBide & Marshman (2006)
- ▷ Frérot & Karagouch (2015)
- ▷ Kiraly (2015)
  
- ▶ Préoccupations d'ordre logistique (budget, disponibilité des salles), compatibilité des logiciels, assistance technique)
  
- ▶ Quels outils choisir (ergonomie vs marché) ? Faut-il privilégier l'apprentissage d'un ou plusieurs outils ? (Maîtrise élargie vs flexibilité ?)
  - ▷ Outils commerciaux dotés d'une fonctionnalité de TA vs TA spécialisé e (moteurs entraînés)
  
- ▶ Quelle formation pour quel profil de traducteur ?
  - ▷ Traducteur indépendant, en entreprise/organisation internationale
  
- ▶ Préoccupation centrale : Comment privilégier l'acquisition de compétences (transversales) du traducteur et non la simple maîtrise technologique d'un outil ?

## Solutions proposées :

- ▶ Faire appel aux outils en open access (accessibilité accrue pour les étudiants)
  - ▷ Moses : Pose la question des compétences informatiques des étudiants et des formateurs ?
- ▶ Utilisation systématique des outils dans les devoirs, projets, etc.
- ▶ Intégration systématique des outils de TAO dans l'ensemble des cours.
  - ▷ Considérations d'ordre éthique.
- ▶ Kenny, Doherty (2014) :
  - Mise en place d'un module renforcé de TA (32h dont 20h de manipulation outillée)
  - Compréhension fine des principes de la TA, utilisation d'un outil en ligne permettant d'entraîner un moteur de TA avec pour objectif de développer l'esprit critique des étudiants.
    - ▷ Considérations d'ordre organisationnelle, disponibilité de formateurs experts.

## Deux grandes absentes:

- ▶ La dimension professionnalisante :
  - ▷ Découverte et gestion des difficultés inhérentes au déploiement d'un nouvel outil dans l'environnement de travail du traducteur (réorganisation du service, investissement financier, opérabilité et niveau de performance, etc.)
  - ▷ « Bring the workplace into the curriculum » (Kiraly 2015)
  
- ▶ La dimension cognitive :
  - ▷ L'automatisation d'une partie du processus de traduction perturbe les stratégies de traduction (démultiplication des efforts cognitifs, nature fragmentée des sorties de TA qui gêne la cohérence textuelle, éloignement du texte source, etc.) (Kübler & Martikainen 2016)
  - ▷ Gestion des émotions des traducteurs face à l'intégration de nouvelles technologies (Koskinen & Ruokonen 2017)



## Projets de recherche appliquée (Peraldi, Clouet 2012) :

- ▶ Projets menés par des étudiants de M1 et M2 pour le compte de véritables commanditaires (entreprises de traduction, organisations internationales, associations humanitaires, centre de recherche, etc.)
- ▶ Ils constituent une initiation à la recherche fondamentale et empirique tout en permettant aux étudiants de se placer dans une démarche de linguistique appliquée
- ▶ Encadrement par un enseignant-chercheur (approche méthodologique) et un référent de l'entreprise/centre de recherche (adéquation du projet/livrables aux besoins du commanditaire)
- ▶ Ils mènent à rédaction d'un mémoire + soutenance en face d'un jury d'enseignants et de professionnels.
- ▶ Approche résolument pragmatique (en permettant aux étudiants de se confronter au réel professionnel)

## Les douze travaux de la TA :

- ▶ Identification des entités nommées (noms propres) (Reverso)
- ▶ Identification des entités nommées (noms géographiques) (Reverso)
- ▶ Constitution de dictionnaires électroniques en vue de l'amélioration d'un moteur de TA linguistique (Reverso)
- ▶ Création d'une interface de post-édition (Systran + ISEP)
- ▶ Comparatif des outils de TAO/TA à destination d'une entreprise de traduction
- ▶ Projet d'annotation/segmentation discursive en vue de l'amélioration d'un outil de reconnaissance de la parole (LIMSI + interprètes en activité)
- ▶ Projet d'annotation sémantique en vue de l'amélioration d'un moteur de traduction statistique (LIMSI)
- ▶ Analyse qualitative des sorties de TA brutes de MT@EC
- ▶ Comparatif biotraduction/TAO + TA pour le compte d'une entreprise de traduction dans le domaine financier
- ▶ Comparatif biotraduction/TAO + TA pour le compte d'une entreprise de traduction dans le domaine juridique
- ▶ Analyse des sorties de TA brutes d'un moteur TA statistique entraîné dans le domaine médical (Cochrane + LIMSI)
- ▶ Post-édition des sorties de TA (Cochrane + LIMSI)

## Exemple de PRA

### Comparatif biotraduction/TA + TAO (Peraldi 2016):

- ▶ Entreprise spécialisée dans la rédaction et traduction d'informations règlementées.
- ▶ Principaux clients: entreprises cotées en Bourse ayant l'obligation de publier chaque année un rapport annuel (AMF).
- ▶ L'entreprise travaille de manière quasi exclusive avec des traducteurs et des réviseurs indépendants, qui constituent un réseau d'environ deux cents personnes.
- ▶ Activité en flux tendu: l'ensemble des rapports annuels doivent publiés en même temps (février)
- ▶ Nécessité de scinder les documents entre une dizaine de traducteurs.
- ▶ Documents : taux de récurrence élevé, terminologie et une phraséologie ultra spécialisée, nombreux chiffres.
- ▶ Impératif : cohérence terminologique et stylistique

## Exemple de PRA

### Approche méthodologique :

- ▶ Initiation aux principes de la traduction automatique statistique
- ▶ État de l'art des méthodes d'évaluation des sorties de TA : évaluations subjectives (typologie d'erreurs) versus métriques automatiques (BLUE, PER, etc.)
- ▶ Réflexion menée autour de la constitution de la typologie d'erreurs (Flanagan 1994) (Vilar, Xu, D'Haro & Ney 2006) (Peraldi, Prost 2011)
- ▶ Échantillonnage et choix des textes à traduire
- ▶ Prise en main de l'outil (calcul du taux de récurrences des segments, intégration de la mémoire de traduction, etc.)

## Typologie d'erreurs

Type d'erreur		Fréquence (%)	
<b>Erreur de terminologie</b>	Nom propre mal traduit		
	Terminologie générale		
	Terminologie financière		
	Terminologie juridique		
	Utilisation de termes qui n'existent pas dans le dictionnaire de la langue concernée		
	Expression idiomatique mal traduite		
	Collocation mal choisie		
<b>Erreur de langue</b>			
	Orthographe	Mot mal orthographié	
		Mauvais emploi des majuscules	
	Grammaire	Erreur de conjugaison	
		Erreur de temps verbal	
		Faute d'accord	
		Absence ou mauvaise utilisation d'un article	
		Absence ou mauvaise utilisation d'une préposition	
		Absence ou mauvaise utilisation d'une conjonction de coordination	
Absence ou mauvaise utilisation d'une apostrophe			
Absence ou mauvais positionnement de la négation			
Syntaxe / Style	Ordre des mots		
	Nécessité de déplacer un mot		

## Catégorisation des erreurs

Exemple 1 :

Texte original en français	Traduction automatique	Traduction en anglais du document de référence	Autres erreurs en cause
<i>Vector s'est porté fort des engagements de souscription pris par Petalite dans l'Offre Améliorée.</i>	<i>Vector is a strong underwriting commitments by Petalite in the improved offer.</i>	<i>Vector has undertaken to stand surety of Petalite's subscription undertakings under the Improved Offer.</i>	<i>Erreur de terminologie générale sur « s'est porté fort »</i>  <i>Erreur de terminologie juridique et financière « underwriting commitments »</i>

Exemple 2 :

Texte original en français	Traduction automatique	Traduction en anglais du document de référence	Autres erreurs en cause
<i>à titre réductible le nombre d'actions nouvelles qu'ils désireraient en sus de celui leur revenant du chef de l'exercice de leurs droits à titre irréductible.</i>	<i>collapsible as the number of new shares that they wish their returning from the head of the exercise of their rights on an irreducible basis over and above that.</i>	<i>on a reducible basis, to the number of new shares they request on top of those attributable to them when exercising their entitlements on an irreducible basis.</i>	<i>Erreur de terminologie juridique : « collapsible as »</i>  <i>Erreur de syntaxe : mauvais ordre des mots « en sus de » est tout à la fin de la phrase</i>  <i>Erreur de grammaire : absence de préposition après « wish »</i>

## Typologie d'erreurs

- ▶ Les segments identiques, bons ou passables : ils sont soit identiques à la traduction de référence, soit jugés de qualité satisfaisante ; ils peuvent être conservés tels quels.
- ▶ Les segments à corriger : ils présentent un style moyen, mais compréhensible ; ils contiennent néanmoins quelques non-sens, erreurs de terminologie et/ou des omissions que le relecteur doit corriger.
- ▶ Les segments à retraduire : ces segments sont incompréhensibles ; le relecteur devra se référer au texte source et retraduire ces segments.

Au 31 décembre 2011, 1.469.580 actions anciennes non regroupées étaient en circulation.	To December 31, 2011, 1.469.580 non consolidated old shares were outstanding.	As at December 31, 2011, 1,469,580 unconsolidated shares remained in circulation.	Langue - grammaire : mauvais choix de prépositions : "to December" DL : non adaptation des nombres dans la langue cible Langue - orthographe : mot mal orthographié - non consolidated Terminologie - terminologie financière : "were outstanding"	Segment à retraduire
Ces actions peuvent être présentées au regroupement jusqu'à l'expiration d'un délai de deux ans à compter du 15 juillet 2010, soit jusqu'au 15 juillet 2012.	These actions may be submitted to the grouping until the expiration of a period of two years from July 15, 2010, ? until July 15, 2012.	Shares may be exchanged for a period of two years with effect from July 15, 2010, i.e., until July 15, 2012.0	DL : Mot non traduit : action Terminologie - terminologie financière : présentées au regroupement DL : omission : "soit"	Segment à corriger
Conformément à l'article L.	In accordance with article L.	In accordance with Article L.		Segment bon / passable

# Résultats

Erreurs de terminologie	Nombre	%
Nom propre mal traduit	28	6,6
Terminologie générale	95	22,5
Terminologie financière	229	54,1
Terminologie juridique	67	15,8
Utilisation de termes qui n'existent pas dans le dictionnaire de la langue concernée	2	0,5
Expression idiomatique mal traduite	0	0,0
Collocation mal choisie	2	0,5
<b>Total</b>	<b>423</b>	<b>31,9%</b>

  

Erreurs de langue	Nombre	%
Orthographe	275	45,3
Grammaire	176	29,0
Syntaxe / Style	141	23,2
Ponctuation	15	2,5
<b>Total</b>	<b>607</b>	<b>45,8%</b>

  

Erreurs de sens	Nombre	%
Faux-sens	24	22,2
Contresens	23	21,2
Non-sens	61	56,4
<b>Total</b>	<b>108</b>	<b>8,2%</b>

  

Dysfonctionnements du logiciel	Nombre	%
Espaces	12	6,4
Mauvaise distinction du segment (liée aux balises...)	0	0,0
Non-reconnaissance des chiffres (dans le texte, numérotation word...)	0	0,0
Non-reconnaissance des dates	5	2,7
Non adaptation à la typographie cible (chiffres)	85	45,5
Mot non traduit/laissé en français	22	11,8
Mot / groupe de mots omis	42	22,5
Passage non traduit	4	2,1
Ajout de mots	11	5,9
Non reconnaissance de la MT de segments déjà existants	0	0,0
Non reconnaissance du singulier/pluriel	6	3,2
<b>Total</b>	<b>187</b>	<b>14,1%</b>

  

<b>TOTAL D'ERREURS</b>	<b>1325</b>
------------------------	-------------

Types de segments	Tous les segments		Sans chiffres	
	quantité	%	quantité	%
Segment identique/ bon/passat	283	42%	125	27%
Segment à corriger	275	41%	216	47%
Segment à retraduire	116	17%	115	25%



## Résultats

	Détail	Coût	Jours
Traduction humaine et relecture	10 156 x 0,12 € = 1218, 72 € 6,06 heures x 35 €/heure = 212 €	1430,72 €	5,83
Post-édition brute	19,82 heures x 35 €/heure	693,70 €	2,5
Post-édition évoluée	22, 13 heures x 35 €/heure	774, 55 €	2,75

Tableau 1 : Estimation des coûts

Tableau 2: comparatif des scores attribués par les juges humains

	Fidélité	Intelligibilité	Style
Document de référence	Traducteur 1 : 9/10	Traducteur 1 : 8/10	Traducteur 1 : 8/10
	Traducteur 2 : 9/10	Traducteur 2 : 6/10	Traducteur 2 : 7/10
Extrait post-édité n° 1	7/10	6/10	6/10
Extrait post-édité n° 2	7/10	8/10	7/10

## Conclusion

### Apports pédagogiques :

- ▶ Mise en œuvre d'une véritable démarche scientifique (état de l'art, méthodologie, résultats et analyse)
- ▶ Découverte et remise en question de nombreux concepts (notion d'évaluation de la TA, stratégies de traduction et de post-édition)
- ▶ Maîtrise syntaxique et phraséologique accrue
- ▶ Découverte de l'activité de post-édition et prise en compte des efforts cognitifs et ajustements nécessaires
- ▶ Prise en main d'un nouvel outil
- ▶ Découverte de l'environnement de travail d'une entreprise
- ▶ Echanges constructifs avec le commanditaire
- ▶ Acquisition de compétences organisationnelles et managériales
- ▶ Collaboration réussie pour le commanditaire qui bénéficie d'un regard neuf et de nouvelles méthodologies de recherche

## Bibliographie

BOWKER Lynn, MCBRIDE Cheryl & MARSHMAN Elisabeth (2008), « Getting more than you paid for? Considerations in integrating free and low-cost technologies into translator training programs », *Redit*, 1, 26-47.

CRONIN, M. (2013). Translation and globalization. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 491–502). London, UK: Routledge.

DEPRAETERE Ilse (2010). What counts as useful advice in a university post-editing training context ? Report on a case study. Viggo Hansen, and François Yvon. Proceedings of the 14th annual conference of the European Association for Machine Translation, 27-28 May 2010, Saint-Raphaël, 2010, available at <http://www.mt-archive.info/EAMT-2010-Depraetere-2.pdf>.

DEPRAETERE Ilse, DE SUTTER Nathalie & TEZCAN Arda (2014), « Post-Edited Quality, Post-Editing Behaviour and Human Evaluation: A Case Study », S. O'Brien, L. Winther-Balling, M. Carl, M. Simard & L. Specia (dir.), *Post-editing of Machine Translation: Processes and Applications*, Newcastle upon Tyne : Cambridge Scholars Publishing, 78-108

ENGESTRÖM, Y., & SANNINO, A., 2010. Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review*, doi:10.1016/j.edurev.2009.12.002.

FREROT C., KARAGOUCHE L. (2016). Outils d'aide à la traduction et formation de traducteurs : vers une adéquation des contenus pédagogiques avec la réalité technologique des traducteurs. *ILCEA Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*, 27 , pp. 120–140

## Bibliographie

DOHERTY, S (2016). The Impact of Translation Technologies on the Process and Product of Translation. In: International Journal of Communication 10(2016), 947–969.

FLANAGAN, M., 1994. Error Classification for MT Evaluation. Proceedings of AMTA 94, 65-72

GASPARI, F., ALMAGHOUT, H., & DOHERTY, S. (2015). A survey of machine translation competences: Insights for translation technology educators and practitioners. Perspectives: Studies in Translatology, 23(3), 333–358

KENNY, D. (2011). The ethics of machine translation. Paper presented at the New Zealand Society of Translators and Interpreters annual conference, Auckland, New Zealand.

KENNY, Dorothy & Stephen DOHERTY (2014). Statistical machine translation in the translation curriculum: overcoming obstacles and empowering translators, The Interpreter and Translator Trainer, 8:2, 276-294.

KIRALY Don (2015), European Graduate Placement Scheme (EGPS) event, en ligne sur <[www.egps.org/6feb2015-egps-day/](http://www.egps.org/6feb2015-egps-day/)>.

KOLETNIK KOROŠEC, Melita (2011). The Internet, Google Translate and Google Translator Toolkit - a nuisance or necessity in translator education? : paper at the Tralogy conference Translation careers and technologies: convergence points for the future, Paris, March 3th & 4th 2011. Paris, 2011

KOSKINEN, K and RUOKONEN, M (2017) ‘Love Letters or Hate Mail? Translators’ technology acceptance in the light of their emotional narratives’. In Kenny, D (ed) Human Issues in Translation Technology, London: Routledge

## Bibliographie

MARTIKAINEN, Hanna et Natalie KÜBLER, « Ergonomie cognitive de la post-édition de traduction automatique : enjeux pour la qualité des traductions », ILCEA [En ligne], 27 | 2016, mis en ligne le 08 novembre 2016, consulté le 11 novembre 2016. URL : <http://ilcea.revues.org/3863>

O'BRIEN, S. (2006). Eye-tracking and translation memory matching. *Perspectives: Studies in Translatology*, 14(3), 185–205.

O'BRIEN, S. (2008). Processing fuzzy matches in translation memory tools—an eye-tracking analysis. In S. Göpferich, A. Jakobsen, & I. Mees (Eds.), *Copenhagen studies in language*, 36 (pp. 79–102). Copenhagen, Denmark: Samfundslitteratur.

PERALDI, Sandrine et Audrey, PROST (2011). « La Post-édition à la portée du traducteur.» Actes de Tralogy 2011. Paris <<http://lodel.irevues.inist.fr/tralogy/index.php?id=107>>

PERALDI, Sandrine et Louis-Marie CLOUET (2012). Articulation de la recherche et de la formation à l'ISIT : un atout crucial face aux problématiques interculturelles actuelles», *Bulletin du. CRATIL*, n°8, janvier 2012, p. 30-35.

PERALDI, Sandrine (2016). De la traduction automatique brute à la post-édition professionnelle évoluée : le cas de la traduction financière. *Revue française de linguistique appliquée*, pp.67-90, 2016.

VILAR David, Jia XU, Luis Fernando D'HARO, and Hermann NEY. 2006. Error analysis of statistical machine translation output. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

# Merci de votre attention !

Des questions ?

▶ [sandrine.peraldi@ucd.ie](mailto:sandrine.peraldi@ucd.ie)



Modèle Powerpoint [SlidesCarnival](#)  
Photos: [Unsplash](#)