

POST-ÉDITION DE TRADUCTION AUTOMATIQUE : QUELS EFFETS SUR LA QUALITÉ DES TEXTES TRAITÉS AU SEIN DE LA DGT ?

RÉSULTATS PRÉLIMINAIRES ET PREMIÈRES TENDANCES

DE FARIA PIRES Loïc

Service de Traduction spécialisée et Terminologie, FTI-EII (**UMONS**)

« Le traducteur à l'ère des nouvelles technologies », ULiège, 6 mai 2019

GENÈSE DU PROJET

Utilisation grandissante des technologies de traduction au sein de la DGT (Commission européenne, 2017 : 3)

Augmentation de la demande en traduction à la DGT (Direction générale de la traduction, 2016 : 4) → Recours croissant à la post-édition de TA (Direction générale de la traduction, 2016 : 10).

→ **Outil à la disposition aussi bien des traducteurs fonctionnaires, temporaires, et des stagiaires.**

HYPOTHÈSES

-La **qualité de la post-édition**, mesurée à l'aune du nombre d'erreurs (ici « modifications ») et de leurs répercussions sur la qualité d'ensemble, n'est pas influencée par le niveau d'expérience des post-éditeurs (Čulo et al., 2014 : 207 ; Daems, 2016 : vii).

-Les **catégories** d'erreurs (ici « modifications » de post-édition) sont, pour leur part, différentes en fonction de l'expérience des post-éditeurs (De Almeida & O'Brien, 2010 : 7 ; Depraetere, 2010 : 1 ; Guerberof Arenas, 2014 : 72 ; Daems, 2016 : viii).

→ **Qu'en est-il dans le contexte institutionnel authentique de la DGT ?**

QUELLE TRADUCTION AUTOMATIQUE À LA DGT ?






- **Jusque début 2019** : TA Statistique MT@EC (Moses) pour la combinaison linguistique EN > FR nous intéressant.
- **Février 2019** : Changement de paradigme → TA neuronale, projet *eTranslation* (moteur *Marian*).

→ Influence du changement de moteur sur la qualité de la PE et la répartition des révisions en catégories ?

ÉTUDE DES SEGMENTS POST-ÉDITÉS

Fondements du projet

- Comparaison des **traducteurs fonctionnaires** et des **stagiaires « Blue Book »**.
- Utilisation de textes **authentiques**, traduits / post-édités à la DGT entre 2017 et 2019-20.
- Étude des **marques de révision dans les segments PE**, ces marques de révision engendrant un coût temporel et budgétaire (Direction générale de la traduction, 2012 : 15).

23	At the same time, we see challenges and we see that the best way to protect the idea of development cooperation and to promote it further is to make our development cooperation more effective than it is today.	 AT	En même temps, nous voyons <u>(sommés conscients)</u> des difficultés. Et nous constatons que la meilleure façon de défendre <u>et de promouvoir</u> l'idée de coopération au développement et de la promouvoir est de rendre notre coopération au développement plus efficace qu'elle ne l'est aujourd'hui.	P
24	We see that the world is changing fast and a lot and we cannot simply continue to do things as they have always been done.	 AT	Nous voyons que le monde change rapidement et beaucoup, et nous ne pouvons pas nous contenter d'agir comme nous l'avons toujours fait.	
25	We have to do it but in the meantime we have the responsibility of thinking about what is coming next.	 AT	Nous devons agir, mais en même temps, nous avons la responsabilité de réfléchir à l'avenir.	
26	In these months and years we have explored in the European Union new ways of doing development cooperation, following the compass of the Sustainable Development Goals (SDGs).	 AT	Ces dernières années et ces derniers mois, nous avons exploré, dans l'Union européenne, de nouvelles façons de faire de la coopération au développement, en gardant comme boussole les objectifs de développement durable (ODD).	P
27	You know that for us partnership with the United Nations and the United Nations system is the compass of all our foreign policy and this is even more true when it comes to development cooperation.	 AT	Vous savez que pour nous, le partenariat avec les Nations unies et le système des Nations unies servent de boussole à l'ensemble de notre politique étrangère, et c'est encore plus vrai en matière de coopération au développement.	

DÉFI DE L'ÉTUDE DE LA QUALITÉ

Précaution méthodologique : Évaluation de la qualité en traduction = inévitablement subjective (Castilho et al. 2018, 30-31) et dépendante du contexte et de la situation de communication propre à chaque texte traduit / post-édité à la DGT (Direction générale de la traduction, 2014 : 1).

- **Objectif** = choisir un outil tenant compte de la visée fonctionnaliste des textes traduits à la DGT (Drugan et al., 2018 : 61) et visant à réduire la subjectivité autant que possible.
- Une fois l'outil choisi, s'assurer de la **fiabilité** de celui-ci.

OUTIL CHOISI

-Outils LISA et DQF préalablement envisagés et écartés.

→ Grille de Toudic et al. (2014) retenue :

9 catégories : 1) Sens / 2) Addition – Omission / 3) Terminologie / 4) Phraséologie / 5) Grammaire – Syntaxe / 6) Orthographe – Typographie / 7) Style / 8) Localisation / 9) PAO.

4 types d'effets : 1) Précision / 2) Fonctionnalité / 3) Lisibilité / 4) Conformité (**Très pertinent dans le cadre des documents très variés post-édités à la DGT**).

Pénalité de 1 à 5 pour chaque marque de révision (à répartir entre max. 2 types d'effets) en fonction de la gravité perçue des éléments ayant poussé le réviseur à intervenir.

Adaptée à la dimension fonctionnaliste de la qualité à la DGT (voir Toudic et al., 2014 : 9).

Limitation des biais subjectifs (Toudic et al., 2014 : 8).

OUTIL CHOISI


-Attribution d'une pénalité à chaque marque de révision dans les contenus post-édités par les volontaires du projet (volet quantitatif) et classement des marques de révisions dans les catégories proposées (volet qualitatif).

-Accord inter-évaluateurs pour vérifier la fiabilité de l'outil (juillet 2018, sur 33 textes, avec une autre chercheure en traductologie) :

→ Pénalités – Kappa de Cohen pondéré linéaire : **0.8919**

→ Catégories – Kappa de Cohen : **0.7377**

TROIS TYPES DE MESURES

- 1) **Nombre moyen de marques de révision** dans les textes post-édités par chaque traducteur (normalisé par tranche de 100 mots pour comparaison).
 - 2) **Pénalité moyenne** attribuée aux marques de révision dans les segments post-édités par chaque volontaire du projet (normalisé /100 mots).
 - 3) **Répartition**, pour chaque traducteur, **des pénalités attribuées aux marques de révision** dans les 9 **catégories** proposées par l'outil de Toudic et al. (2014).
- 

TRADUCTEURS ET NOMBRE DE TEXTES/MOTS PE

Traducteur	Nombre de textes	Nombre de mots post-édités
Expérimenté 1	18	24 649
Expérimenté 2	8	4 813
Expérimenté 3	9	12 074
Expérimenté 4	4	5 642
Expérimenté 5	49	36 776
Peu expérimenté 1	7	11 822
Total	95	95 776

Corpus provisoire, dont la représentativité sera améliorée au cours de l'année.

EXEMPLE D'ÉVALUATION ET DE CLASSEMENT

33	This summary of the scientific Opinion on 'Potential risks to human health of Light Emitting Diodes (LEDs)' by the Scientific Committee on Health, Environmental and Emerging Risks (SCHEER) covers some of the Opinion's key points and goes more in depth than the one-page factsheet on the same topic, also available on this website.	AT	Cette synthèse de l'avis scientifique du CSRSEE sur les risques potentiels pour la santé humaine des diodes électroluminescentes (LED) émis par le comité scientifique des risques sanitaires, environnementaux et émergents (CSRSEE) couvre certains points essentiels de l'avis et va plus loin que la fiche d'information d'une page sur le même sujet, également disponible sur ce site.	P+
34	An abstract and a shorter but more technical summary are also included in the scientific Opinion itself .	AT	L' avis scientifique contient également un abrégé et un résumé plus bref mais plus technique.	
35	Information about data and methodology and the science behind LEDs and eye and skin optics are also found in the Opinion and are not covered here.	AT	Des informations relatives aux données, à la méthodologie et aux aspects scientifiques des LED et des effets des rayonnements optiques sur la peau et les yeux figurent également dans l'avis, mais ne sont pas -reprises ici.	
36	Navigation	99%	Navigation	H+
37	1.	100%	1.	LI+
38	Background	99%	Contexte	

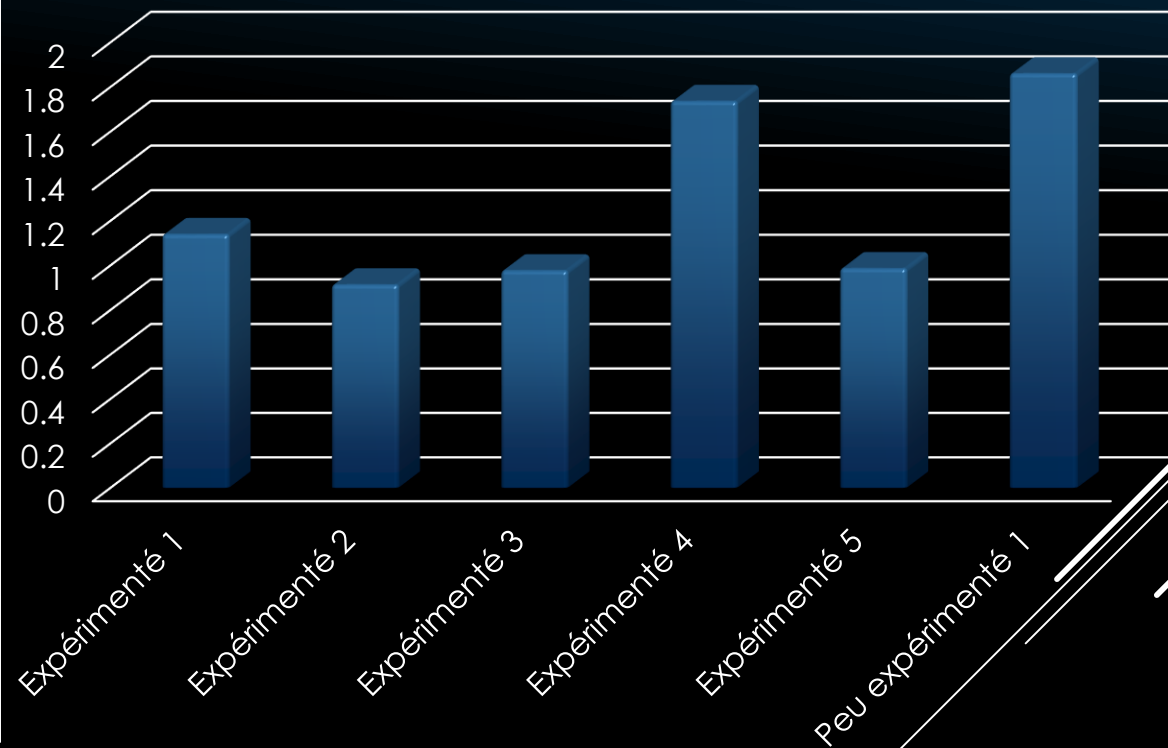
Numéro	Numéro du segment	Texte source	TA brute	Post-édition	Révision	Type de modification	Modification précise	Précision	Fonctionnalité	Lisibilité	Conformité	Pénalité totale/erreur
1	16	traffic signals and increasingly in the automotive industry, to name just some of the applications.	les signaux de trafic et, de plus en plus, l'industrie automobile, pour n'en citer que quelques-unes.	les feux de circulation et, de plus en plus, l'industrie automobile.	les feux de signalisation et, de plus en plus, l'industrie automobile.	Terminologie	Variante inappropriée	2				2
2	33	This summary of the scientific Opinion on 'Potential risks to human health of Light Emitting Diodes (LEDs)' by the Scientific Committee on Health, Environmental and Emerging Risks (SCHEER) covers some of the Opinion's key points and goes more in depth than the one-page factsheet on the same topic, also available on this website.	Le présent résumé de l'avis scientifique sur les «risques potentiels pour la santé humaine des diodes électroluminescentes (DEL)» par le comité scientifique des risques sanitaires, environnementaux et émergents (CSRSEE) couvre certains points essentiels de l'avis et va plus loin que la fiche d'information d'une page sur le même sujet, également disponible sur ce site.	Cette synthèse de l'avis scientifique sur les risques potentiels pour la santé humaine des diodes électroluminescentes (LED) émis par le comité scientifique des risques sanitaires, environnementaux et émergents (CSRSEE) couvre certains points essentiels de l'avis et va plus loin que la fiche d'information d'une page sur le même sujet, également disponible sur ce site.	Cette synthèse de l'avis scientifique du CSRSEE sur les risques potentiels pour la santé humaine des diodes électroluminescentes (LED) couvre certains points essentiels de l'avis et va plus loin que la fiche d'information d'une page sur le même sujet, également disponible sur ce site.	Style	Manque de fluidité			1		1

Comparaison possible avec la TA brute

RÉSULTATS PARTIELS : MOYENNE DES MARQUES DE RÉVISION

Traducteur	Nombre moyen de modifications (/100 mots)	n modifications	n mots post-édités
Expérimenté 1	1,1400	281	24 649
Expérimenté 2	0,9141	44	4 813
Expérimenté 3	0,9773	118	12 074
Expérimenté 4	1,7369	98	5 642
Expérimenté 5	0,9870	363	36 776
Peu expérimenté 1	1,8609	220	11 822

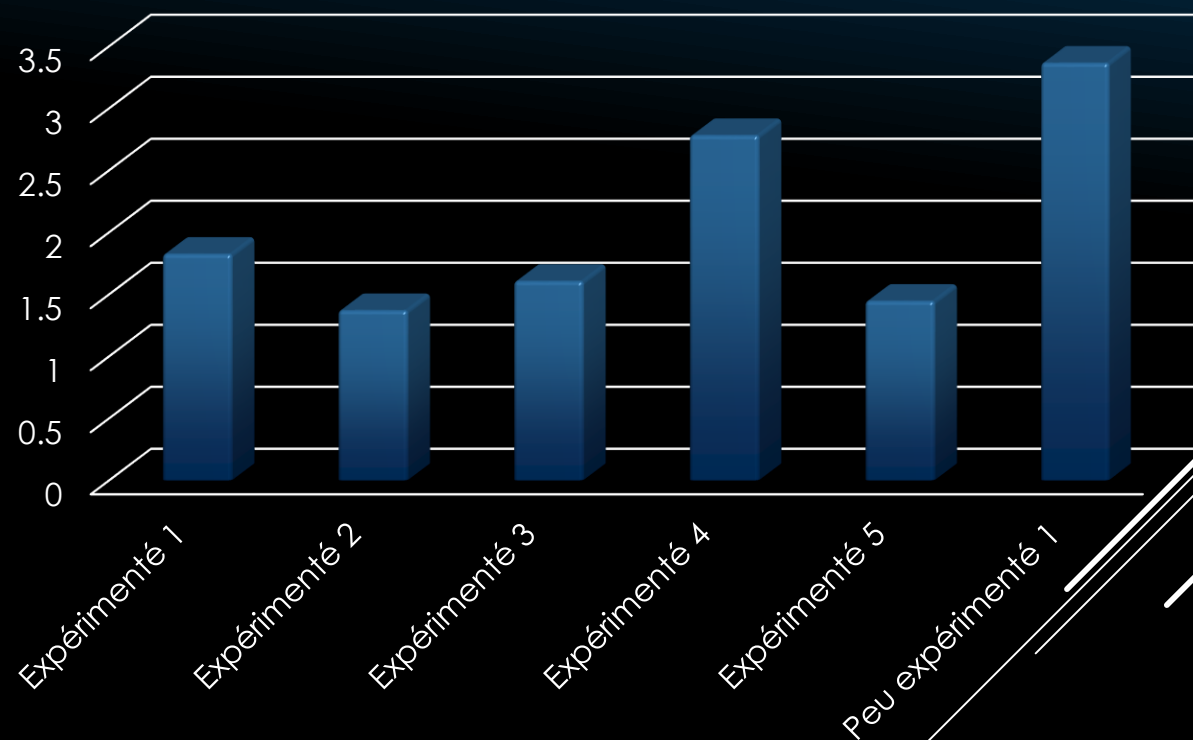
Nombre moyen de marques de révision /100 mots



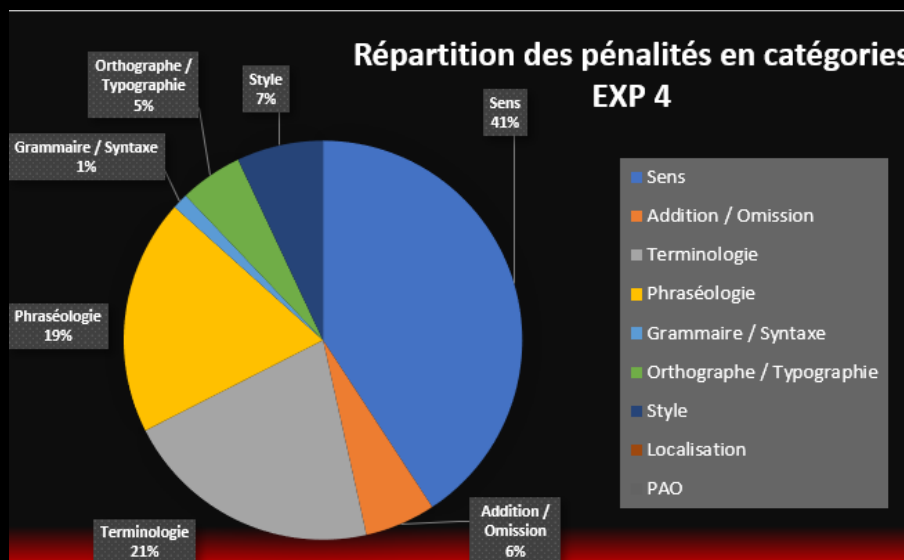
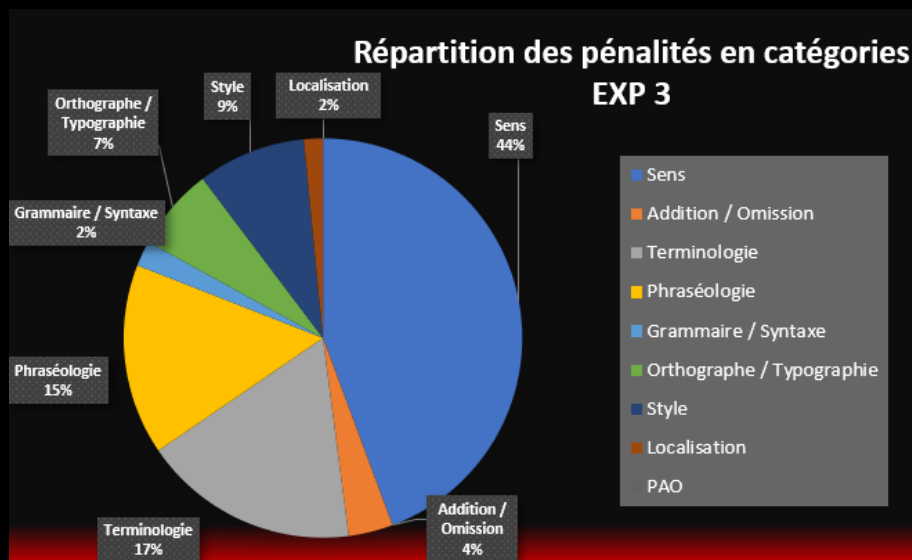
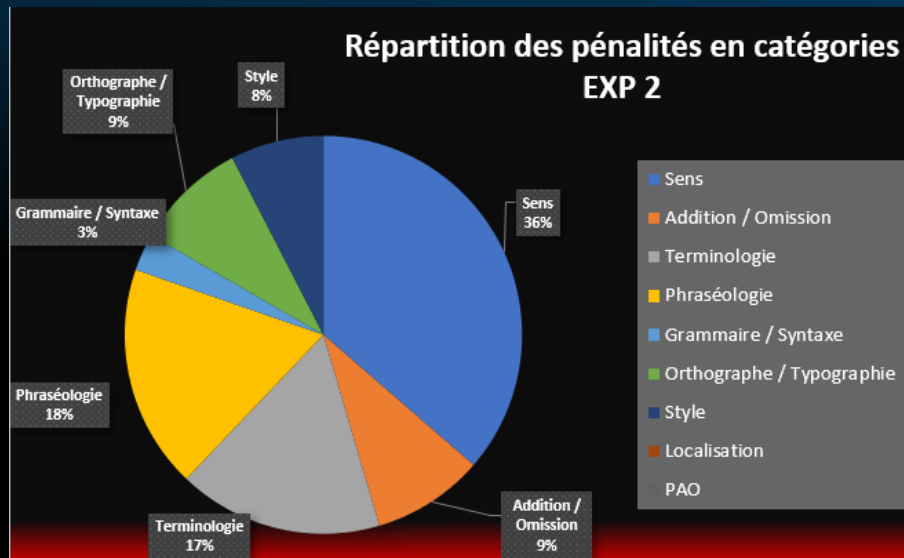
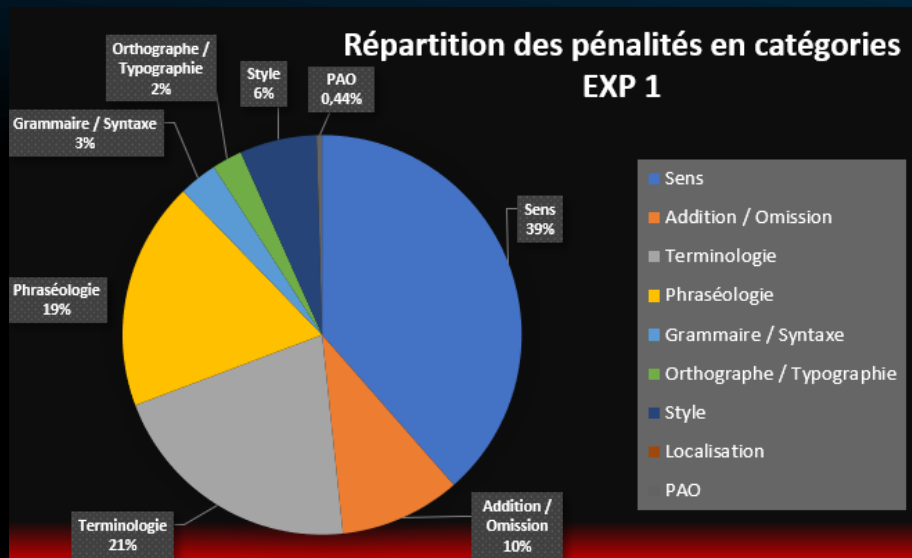
RÉSULTATS PARTIELS : MOYENNE DES PÉNALITÉS

Traducteur	Pénalité moyenne (/100 mots)	n pénalités	n mots post-édités
Expérimenté 1	1,8256	450	24 649
Expérimenté 2	1,3713	66	4 813
Expérimenté 3	1,6067	194	12 074
Expérimenté 4	2,7827	157	5 642
Expérimenté 5	1,4466	532	36 776
Peu expérimenté 1	3,3666	398	11 822

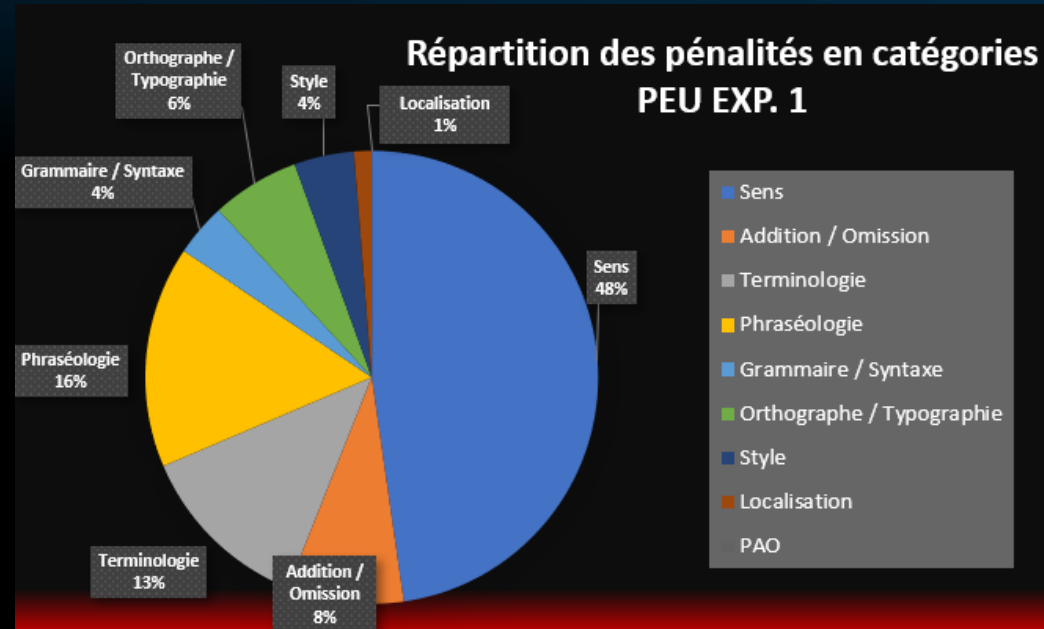
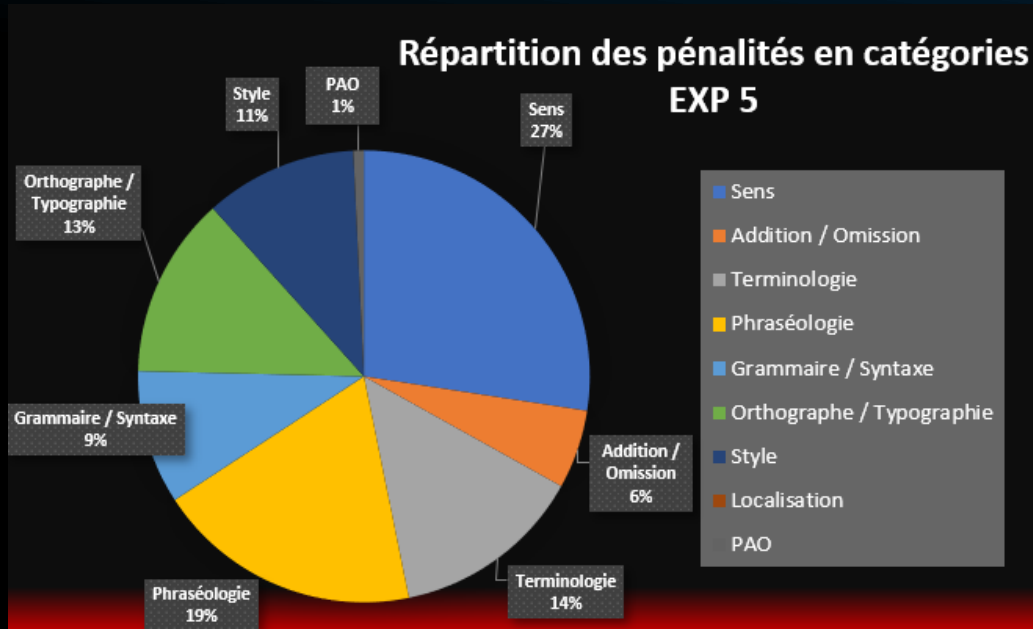
Pénalité moyenne /100 mots



RÉSULTATS : CATÉGORIES DE MODIFICATIONS



RÉSULTATS : CATÉGORIES DE MODIFICATIONS



CONCLUSIONS PROVISOIRES (% NB MARQUES DE RÉVISION)

Rappel : « Le niveau d'expérience n'influence pas la qualité de la post-édition ».

Selon le nombre de marques de révision... hypothèse **non vérifiée** :

-Traducteurs EXP 1, 2, 3 et 5 = résultats plutôt homogènes.

MAIS

-Traducteur P. EXP 1 = quantité de révisions la plus élevée.

ET

-Traducteur EXP 4 = quantité de marques de révision élevée par rapport aux autres EXP.

→ **Quelques variations individuelles observées, faible expérience pourrait (?) nuire à la qualité de la PE**

CONCLUSIONS PROVISOIRES (% PÉNALITÉS)

Rappel : « Le niveau d'expérience n'influence pas la qualité de la post-édition ».

Selon les pénalités attribuées aux marques de révision... hypothèse **non vérifiée** :

- Traducteurs EXP 2 et 5 = pénalités moyennes les plus basses.
- Traducteurs EXP 1 et 3 = pénalités intermédiaires.
- Traducteurs EXP 4 et PEU EXP 1 = pénalités les plus élevées.

→ Variations individuelles encore plus marquées (affinités individuelles avec la TA ? / Différents niveaux de tolérance des réviseurs ? → Focus Group à venir pour le déterminer)

CONCLUSIONS PROVISOIRES (% CATÉGORIES)

Rappel : « Les catégories d'erreurs [des modifications des réviseurs] sont différentes en fonction de l'expérience des post-éditeurs ».

-Guerberof Arenas (2014 : 72) : Peu expérimentés = problèmes de terminologie majoritaires / Expérimentés : problèmes de langue et style majoritaires.

-Depraetere (2010 : 1) : Problèmes récurrents chez les traducteurs peu expérimentés pratiquant la PE = calques et pertes de sens.

-Daems (2016 : viii) : Davantage de problèmes de sens chez les traducteurs peu expérimentés que chez les expérimentés lors de la PE.

CONCLUSIONS PARTIELLES (% CATÉGORIES)

Rappel : « Les catégories d'erreurs sont différentes en fonction de l'expérience des post-éditeurs ».

Hypothèse générale **partiellement vérifiée** :

-Hypothèse de Guerberof Arenas (2014 : 72) : Peu vérifiée. Peu expérimenté 1 affiche la plus faible proportion de pénalités en termes de terminologie. Certes, il affiche la plus faible proportion de pénalités relatives au style, mais le reste des aspects linguistiques (phraséologie, grammaire / syntaxe...) obtiennent une proportion de pénalités similaire aux traducteurs expérimentés.

-Hypothèse de Depraetere (2010 : 1) : Vérifiée (sens). Davantage de pénalités attribuées au sens (sens + addition / omission, pour un total de 56 % des pénalités) chez le traducteur peu expérimenté 1. Taux supérieur à tous les traducteurs expérimentés. *Calques : trop peu pour vérifier cet aspect.*

-Hypothèse de Daems (2016 : viii) : Vérifiée. Même constat : pénalités liées au sens plus récurrentes chez le traducteur peu expérimenté que chez les autres.

→ Niveau d'expérience semble influencer les catégories de révisions, MAIS des variations individuelles apparaissent entre les traducteurs expérimentés.

LIMITES PROVISOIRES ET PISTES DE SOLUTIONS

→ **Manque de données de traducteurs peu expérimentés.**

- Collecte de davantage de textes en cours.

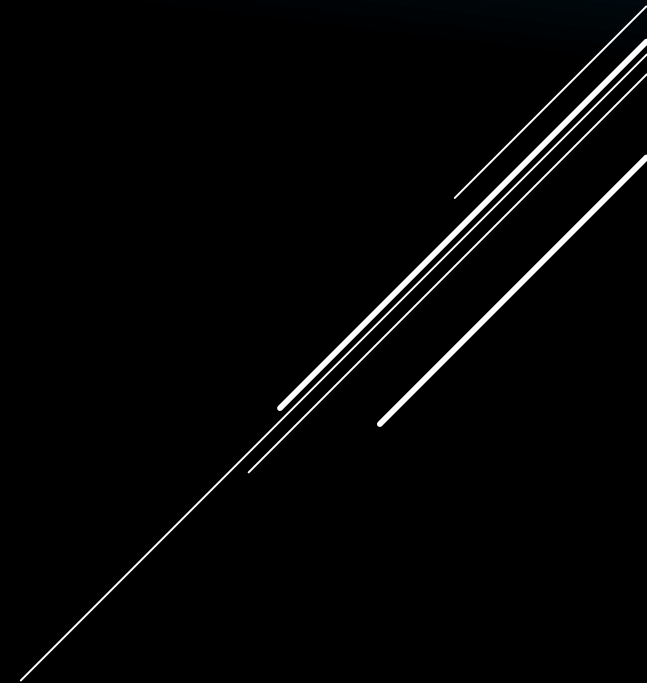
→ **Quantité de données inégales pour les traducteurs expérimentés (contrainte inhérente au contexte institutionnel authentique de cette recherche et au caractère volontaire de la participation...).**

- Dialogue avec les responsables de la DGT en cours pour équilibrer les données.
- 

DISCUSSION

Merci de votre attention !

Loic.DEFARIAPIRES@umons.ac.be



BIBLIOGRAPHIE

- Castilho, S., Doherty, S., Gaspari, F., & Moorkens, J. (2018). Approaches to Human and Machine Translation Quality Assessment. In J. Moorkens, Castilho, S., Gaspari, F., & Doherty, S. (Ed.), *Translation Quality Assessment - From Principles to Practice (Vol. 1, pp. 9-38)*. Cham, Switzerland: Springer.
- Commission européenne. (2017). *Outils d'aide à la traduction et cycle de travail*. Luxembourg: Office des publications de l'Union européenne.
- Čulo, O., Gutermuth, S., Hansen-Schirra, S. & Nitzke, J. (2014). The Influence of Post-Editing on Translation Strategies *Post-editing of Machine Translation: Processes and applications (pp. 200-218)*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Daems, J. (2016). *A translation robot for each translator? - A comparative study of manual translation and post-editing of machine translations: process, quality and translator attitude. (Doctor in de vertaalwetenschap)*, Universiteit Gent, Gent.
- De Almeida, G., & O'Brien, S. . (2010). *Analysing post-editing performance: correlations with years of translation experience. Proceedings of the 14th Annual Conference of the European Association for Machine Translation, Saint-Raphael, France.*
- Depraetere, I. (2010). *What counts as useful advice in a university post-editing training context? Report on a case study. Paper presented at the 14th Annual EAMT Conference, Saint-Raphael.*
- Direction générale de la traduction. (2012). *Quantifying Quality Costs and the Cost of Poor Quality in Translation*. Luxembourg: Office des publications de l'Union européenne.
- Direction générale de la traduction. (2016). *Strategic Plan 2016-2020*.
- Drugan, J., Strandvik, I. & Vuorinen, E. (2018). Translation Quality, Quality Management and Agency: Principles and Practice in the European Union Institutions. In J. Moorkens, Castilho, S., Gaspari, F., & Doherty, S. (Ed.), *Translation Quality Assessment - From Principles to Practice (pp. 39-68)*: Springer.
- Guerberof Arenas, A. (2014). The Role of Professional Experience in Post-editing from a Quality and Productivity Perspective. In S. O'Brien, Winther Balling, L., Carl, M., Simard, M. & Specia, L. (Ed.), *Post-editing of Machine Translation: Processes and Applications*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Toudic, D., Hernandez Morin, K., Moreau, F., Barbin, F., & Phuez, G. (2014). Du contexte didactique aux pratiques professionnelles : proposition d'une grille multicritères pour l'évaluation de la qualité en traduction spécialisée. *ILCEA. Revue de l'Institut des langues et cultures d'Europe, Amérique, Afrique, Asie et Australie*(19). Récupéré à l'adresse <https://ilcea.revues.org/2517>