
Ultrad: outil de traduction de La Presse Canadienne

— La traduction automatique dans
un contexte journalistique —

Traduction et journalisme au Québec

Depuis la fin du 19^e siècle, le journalisme québécois traduit des sources canadiennes, britanniques, américaines.

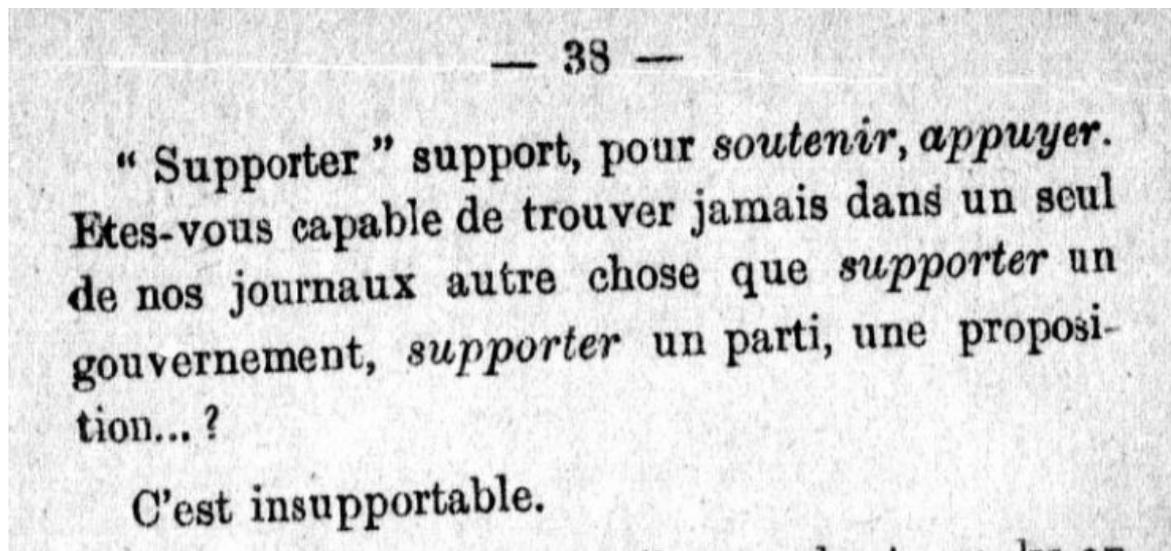
l'ensemble de la presse quotidienne, on s'éloigne de plus en plus de la situation qui prévalait dans plusieurs journaux politiques à la fin du XIX^e siècle. Thomas Chapais évoque ainsi cette époque :

Lorsque nous sommes entré dans la presse, un rédacteur en chef était un homme à tout faire. Aidé à peine d'un ou deux sous-rédacteurs, il devait mettre la main aux différentes besognes qui ont pour objectif d'emplir les colonnes du journal : traduction des dépêches, rédaction des faits divers, des comptes rendus, voire des annonces et des réclames, correction des épreuves, et même surveillance de la mise en pages. Il lui fallait souvent, pour un même numéro, s'occuper de tout cela en même temps [...] Le rédacteur en chef jouait le rôle de factotum, au grand et lamentable détriment de la rédaction. On comprend à quel labeur acharné il était tenu,

Thomas Chapais (1905). *Mélanges de polémiques et d'études religieuses, poitiques et littéraires*. L'Événement, Québec; cité dans Jean de Bonville (1988). *La presse québécoise de 1884 à 1914: genèse d'un média de masse*. PUL, Québec.

Traduction et journalisme au Québec

Il est reconnu que ces traductions effectuées par des journalistes (dont le métier n'est pas la traduction) sont de **qualité discutable**.



Omer Héroux (1888). *Anglicismes et canadianismes*. C. Darveau, Québec.

La Presse Canadienne

Fondée le 17 septembre 1917. Près de 105 ans d'histoire!

En 1959, embauche son premier journaliste francophone.

Donc, entre 1917 et 1959, « *CP [Canadian Press] was an English-language institution, headquartered in Toronto, that operated almost entirely in English. there was no original reporting in French; the Service français was essentially an **improved translation service.*** »

Gene Allen (2013). *Making National News*. University of Toronto Press, Toronto.
Emphase nôtre.

La Presse Canadienne

Aujourd'hui, la PC compte 180 journalistes partout au Canada, dont plus de 40 à Montréal.

Plusieurs journalistes traduisent quotidiennement en français :

- les dépêches de la Canadian Press
- les dépêches de l'Associated Press (dont la PC est le distributeur exclusif au Canada)



En 2018, outil créé pour accélérer cette traduction. Premier outil de TAO utilisé par PC.

Basé sur l'API Cloud Translation de Google (sans entraînement propre à la PC)





TWITTER-MUSK

Assigné à

CPG

Révisé, prêt à exporter ▾

Notes ≡

Outils ↗

Sauver 🗄

Lexiques

Affaires

Version d'origine (EN) (Lecture seule) ▾

Tesla CEO Elon Musk is offering to buy Twitter, saying the social media platform he has criticized for not living up to free speech principles needs to be transformed as a private company.

Twitter Inc. said in a regulatory filing on Thursday that Musk, currently the company's biggest shareholder, has proposed buying the remaining shares of Twitter that he doesn't already own at \$54.20 per share, an offer worth more than \$43 billion.

Musk called that price his best and final offer, although he provided no details on financing. The offer is non-binding and subject to financing and other conditions.

"I invested in Twitter as I believe in its potential to be the platform for free speech around the globe, and I believe free speech is a societal imperative for a functioning democracy," Musk said in the filing. "However, since making my investment I now realize the company will neither thrive nor serve this societal imperative in its current form. Twitter needs to be transformed as a private company."

Twitter shares rose to \$47.83, up 4.3% but well below Musk's offer price, a sign that some investors may doubt the deal will go through. The stock is still down from its 52-week high of about \$73.

Traduction automatique (FR) (Lecture seule) ▾

Le PDG de Tesla, Elon Musk, propose d'acheter Twitter, affirmant que la plate-forme de médias sociaux qu'il a critiquée pour ne pas respecter les principes de la liberté d'expression doit être transformée en une entreprise privée.

Twitter Inc. a déclaré jeudi dans un dossier réglementaire que Musk, actuellement le principal actionnaire de la société, a proposé d'acheter les actions restantes de Twitter qu'il ne possède pas déjà à 54,20 dollars par action, une offre d'une valeur de plus de 43 milliards de dollars.

Musk a qualifié ce prix de sa meilleure et dernière offre, bien qu'il n'ait fourni aucun détail sur le financement. L'offre est sans engagement et soumise à des conditions de financement et autres.

"J'ai investi dans Twitter car je crois en son potentiel d'être la plate-forme de la liberté d'expression dans le monde entier, et je pense que la liberté d'expression est un impératif sociétal pour une démocratie qui fonctionne", a déclaré Musk dans le dossier. "Cependant, depuis que j'ai fait mon investissement, je réalise maintenant que l'entreprise ne prospérera ni ne servira cet impératif sociétal dans sa forme actuelle. Twitter doit être transformé en une entreprise privée.

← TWITTER-MUSK

Assigné à

CPG

Révisé, prêt à exporter ▾

Notes ≡

Outils ↗

Sauver 🗑

Lexiques **Affaires**

Version d'origine (EN) (Lecture seule) ▾

Tesla CEO Elon Musk is offering to buy Twitter, saying the social media platform he has criticized for not living up to free speech principles needs to be transformed as a private company.

Traduction automatique (FR) (Lecture seule) ▾

Le PDG de Tesla, Elon Musk, propose d'acheter Twitter, affirmant que la plate-forme de médias sociaux qu'il a critiquée pour ne pas respecter les principes de la liberté d'expression doit être transformée en une entreprise privée.

On peut afficher, dans les volets de gauche et de droite, l'une ou l'autre des trois versions possibles d'un même texte:

- Version d'origine (EN)
- Traduction automatique (FR)
- Traduction manuelle (FR)

form. Twitter needs to be transformed as a private company.”

Twitter shares rose to \$47.83, up 4.3% but well below Musk's offer price, a sign that some investors may doubt the deal will go through. The stock is still down from its 52-week high of about \$73.

impératif sociétal pour une démocratie qui fonctionne”, a déclaré Musk dans le dossier. "Cependant, depuis que j'ai fait mon investissement, je réalise maintenant que l'entreprise ne prospérera ni ne servira cet impératif sociétal dans sa forme actuelle. Twitter doit être transformé en une entreprise privée.

État de l'art en post-édition - utilité et qualité

Avantages :

- Plus grande **efficacité/rapidité** de la PE que de la traduction; Peraldi (2016), Jia et autres (2019), Wnuk et Wol (2021); confirmé par d'autres études quantitatives (mais l'évaluation de la qualité se limite à la correction linguistique ou est prise en compte de manière holistique)
- Chez les étudiants : Impact inverse chez les étudiants faibles (amélioration) et les étudiants forts (réduction)
- Évitement des calques fautifs, des régionalismes et des fautes de grammaire, Schumacher (2019).

État de l'art en post-édition - utilité et qualité

Inconvénients :

- Tendance à ne pas remettre en question les propositions de la TA -> *priming effect* de Carl et Schaeffer (2017) cité par Volkart et autres (2021), équivalence formelle de Martikainen et Kübler (2016) ou *hallucinations* de Mion (2022), nouvelle déverbalisation à intégrer en cible;
- Reproduction des calques fautifs, effacement d'information (rare), problèmes de référent et de biais de genres, incohérences temporelles et irrégularités terminologiques, Schumacher (2019);
- Effet inattendu (normal en révision): prédominance de la cible en post-édition, Volkart et autres (2021).

Question de recherche

La TA est-elle appropriée pour répondre aux besoins d'une organisation journalistique comme la Presse Canadienne?

Notamment sur le plan de la **qualité** des traductions.

Méthodologie

Bassin initial de **2122** textes



Ultrad

Muslims-Celebrate-Ramadan-révisé

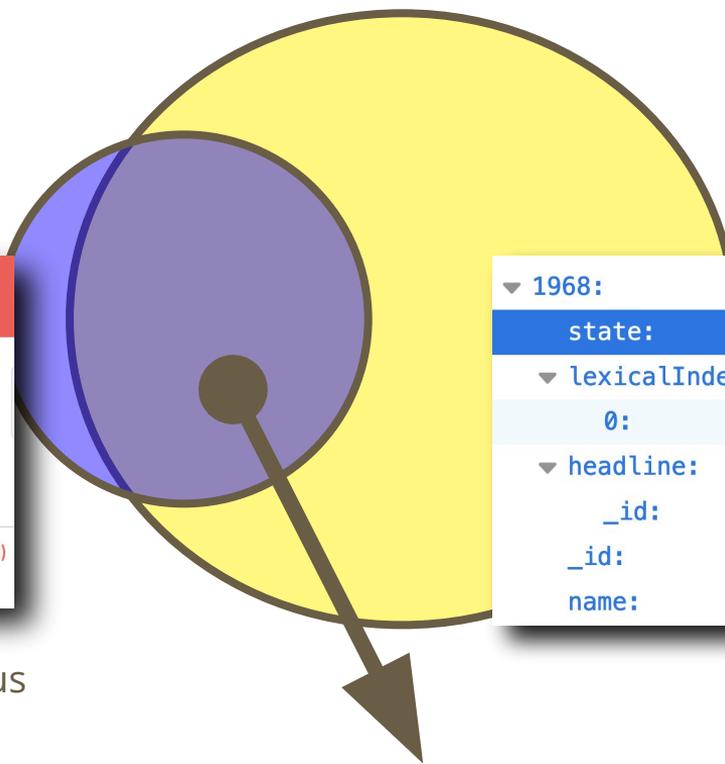
Lexiques Général

Version d'origine (EN)

Canadian Muslims ready to celebrate Ramadan

178 textes étiquetés pour nous par des artisans de la PC en ajoutant «REVISÉ» dans leur titre

624 dont l'état est «revised».



```
▼ 1968:
  state: "revised"
  ▼ lexicalIndexes:
    0: "5b2a6180481dc00014b3a64c"
  ▼ headline:
    _id: "62058ae091fb8800148dad81"
    _id: "62058ae091fb8800148dad82"
    name: "CAMIONS-OTTAWA"
```

Échantillon final de 148 textes.

Méthodologie

3 versions de chaque texte:

- source (EN)
- machine (FR)
- humain (FR)

TH subit d'autres traitements :

- Antidote
- Révision par pupitreur

```
▼ 227:
  state: "revised"
  ▶ lexicalIndexes: [...]
  ▶ headline: {...}
  _id: "6248f18d46e5990014d803c7"
  name: "DANIEL-CRAIG-COVID-RÉVISÉ"
  ▼ text:
    _id: "6248f18d46e5990014d803c8"
    ▶ original: "Could it be the curse of...kman and Sutton Foster."
    ▶ machine: "Serait-ce la malédiction...ckman et Sutton Foster."
    ▶ edited: "Serait-ce la malédiction...ckman et Sutton Foster."
  notes: ...
  updatedAt: "2022-04-03T02:02:13.192Z"
  ▶ lang: {...}
  createdAt: "2022-04-03T00:59:57.274Z"
  __v: 0
  assignedTo: "c50UjoG95tUeszPXdyf3NBPjp3c2"
```

Méthodologie - alignement de segments source et cible

Deux méthodes/logiciels d'alignement : *LF Aligner* et *Logiterm*

- **Six groupes de fichiers alignés**
- Deux fois trois corpus de 148 fichiers alignés : TS-TA (ang-fr), TS-TH (ang-fr) et TA-TH (fr-fr)
- **Paire de segments** = Logiciel aligne une ou plus. phrases sources avec une ou plus. phrases cibles.
- **Sous-segmentation des segments en phrase** = « parseur » maison en Python. Chaque paire de segments source et cible est caractérisée par le nombre de phrases qu'elle contient : 0, 1, 2, ou plus.

Méthodologie - analyse intersegmentale

Typologie de paires de segments (5)

- Paire asymétrique à segment zéro (0-1, 1-0)
- paire symétrique simple (1-1),
- paire symétrique multiple (2-2, 3-3, etc.),
- paire asymétrique simple (1-2, 2-1, 2-3, 3-2, etc.),
- paire asymétrique multiple (1-3, 3-1, 2-4, 4-2, etc.)

~85% sont des paires symétriques simples

Méthodologie - analyse intersegmentale

Indices de qualité de l'alignement

- Nombre de paires totales (plus il est élevé, plus l'analyse est fine)
- Nombre de paires à segment zéro asymétrique (plus il est élevé, moins il y a de faux alignements asymétriques mais plus il y a d'erreurs d'alignement = segments orphelins)
- Nombre de paires asymétriques simples : format asymétrique de base (2-1 ou 1-2); alignement de qualité si le nombre de caractères est égal
- Nombre de paires asymétriques multiples ($\text{diff} > +1$) : plus il est élevé moins l'alignement est de qualité; si l'alignement n'est pas fautif, le format asymétrique est souvent réductible à une forme simple de type 2-1 ou 1-2

Méthodologie - analyse « sac de mots »

Comparaisons statistique « *bag of words* » entre les deux VF:

- termes (non-lemmatisés) les plus utilisés
- mots lexicaux (lemmatisés) les plus utilisés (sauf adverbes)

Résultats

Plusieurs cas où la TH améliore la TA.

Certains cas où la TH ne semble pas avoir amélioré la TA.

Résultats : amélioration normative (norme de rédaction)

Ajouts

(repérés par approche «sac de mots»)

Terme	machine	humain	%humain
Mme	0	78	100,0%
M.	2	241	99,2%

source_EN	machine_FR	humain_FR
Lich was arrested Feb. 17	Lich a été arrêté le 17 février	Mme Lich avait été arrêtée le 17 février, à Ottawa

Résultats : amélioration normative (précision)

Modifications sans décalage
(repérées par approche «sac
de mots»)

Terme	machine	humain	%humain
mercredi	24	52	68,4%
mardi	30	59	62,0%
jeudi	32	51	61,5%
vendredi	40	55	57,9%

source_EN	machine_FR	humain_FR
At a press conference today organized by the NDP	Lors d'une conférence de presse organisée aujourd'hui par le NPD	Lors d'une conférence de presse organisée mercredi par le NPD

Résultats : amélioration normative (mot juste)

Modifications sans décalage
(repérées par approche «sac
de mots»)

Norme journalistique anglo-saxonne:
citations toujours avec le verbe
to say.

Terme	machine	humain	%humain
indiquer	22	68	75,6%
soutenir	25	55	68,8%
affirmer	21	39	65,0%

source_EN	machine_FR	humain_FR
"What you're about to hear will be distressing, [...]" Payne said .	"Ce que vous êtes sur le point d'entendre sera affligeant, [...]", a déclaré Payne.	«Ce que vous êtes sur le point d'entendre sera affligeant, [...]», a prévenu Mme Payne.

Résultats : amélioration normative (mot juste)

Modifications avec décalage
(repérées par approche «sac
de mots»)

Terme	machine	humain	%humain
province	67	43	39,1%

source_EN	machine_FR	humain_FR
The province says the more than \$5-billion investment...	La province affirme que l'investissement de plus de 5 milliards de dollars...	Le gouvernement ontarien affirme que l'investissement de plus de 5 milliards \$...

Résultats : amélioration stylistique intersegmentale

Paires 1-1 -> 1-2 asymétriques

Repéré par sous-segmentation
des phrases - décalage

source	machine	humain
Subordination. Une phrase.	Subordination. Une phrase.	Coordination. Deux phrases.

source_EN	machine_FR	humain_FR
President Joe Biden on Saturday called for Russian leader Vladimir Putin's removal, saying , "For God's sake, this man cannot remain in power."	Le président Joe Biden a appelé samedi à la destitution du dirigeant russe Vladimir Poutine, déclarant : "Pour l'amour de Dieu, cet homme ne peut pas rester au pouvoir".	Le président Joe Biden a appelé samedi à la destitution du dirigeant russe Vladimir Poutine. «Pour l'amour de Dieu, cet homme ne peut pas rester au pouvoir», a lancé le dirigeant américain dans un discours prononcé à Varsovie, en Pologne.

Résultats : contraintes énonciatives (calque corrigé)

Paires hétéromorphes TH

Ratio PIT positif

Servitude du français

source	machine	humain
-	Traductions littérales	Précisions référentielles

source_EN	machine_FR	humain_FR
<p>The consultancy estimates that 140,460 new light vehicles were sold in the month, down from a year ago but up from nearly 100,000 in February.</p>	<p>Le cabinet de conseil estime que 140 460 véhicules légers neufs ont été vendus au cours du mois, en baisse par rapport à il y a un an, mais en hausse par rapport à près de 100 000 en février.</p>	<p>Environ 140 460 véhicules légers neufs ont été vendus au cours du mois, a estimé DesRosiers, un chiffre en baisse par rapport à celui du même mois l'an dernier, mais en hausse par rapport aux près de 100 000 véhicules écoulés en février.</p>

Résultats : traduction littérale abusive

Distance euclidienne TH
3 fois la DE moyenne
caractères, mots et VIN

source	machine	humain
-	Traduction littérale	Non sens corrigé

source_EN	machine_FR	humain_FR
And now most mandates are lifting.	Et maintenant, la plupart des mandats sont levés.	Et maintenant, l'obligation de le porter est levée à plusieurs endroits au pays.

Résultats : amélioration linguistique intrasegmentale

Paires isomorphes TA-TH
Détection «manuelle» par
comparaison lexicale

source	machine	humain
-	Traductions littérales abusives	Mots justes

source_EN	machine_FR	humain_FR
Other high- profile shows on Broadway have had to cancel shows due to their lead actor's contracting the virus, including both "The Music Man" leads, Hugh Jackman and Sutton Foster.	D'autres émissions de haut niveau à Broadway ont dû annuler des émissions en raison de la contraction du virus par leur acteur principal, y compris les deux protagonistes de "The Music Man", Hugh Jackman et Sutton Foster.	D'autres spectacles de grande notoriété à Broadway ont dû annuler leurs représentations en raison de la propagation du virus parmi leurs acteurs principaux, y compris les deux protagonistes de «The Music Man», Hugh Jackman et Sutton Foster.

Résultats : hallucination, corrigée par TH

Modifications : correction du sens (ratio PIN hors norme)

Contresens

source	machine	humain
-	Contresens	Rétablissement du sens

source_EN	machine_FR	humain_FR
Lich was arrested Feb. 17 and initially denied bail , but that decision was reversed March 7 following a review.	Lich a été arrêté le 17 février et a d'abord refusé la mise en liberté sous caution , mais cette décision a été annulée le 7 mars à la suite d'un examen.	Mme Lich avait été arrêtée le 17 février, à Ottawa. Sa demande de remise en liberté sous caution avait d'abord été rejetée , mais cette décision a été annulée le 7 mars à la suite d'un examen.

Résultats

Exemples où la TH n'a pas amélioré la TA

Résultats : calque fautif non corrigé par TH

Antidote

Analyse des détections

Illustration du *priming effect*

source	machine	humain
Terme juridique	Calque et faux sens	Hallucination - Calque fautif

source_EN	machine_FR	humain_FR
Bill C-11 updates sections of a previous bill	Le projet de loi C-11 met à jour des sections d'un projet de loi précédent	Le projet de loi C-11 met à jour des sections d'un projet de loi précédent

Résultats : hypercorrection intrasegmentale

Paires hétéromorphes TA-TH

Ratio PIT positif

Hypercorrection

source	machine	humain
Explicitation	Traductions littérales correctes	Pléonasme

source_EN	machine_FR	humain_FR
About 3 inches (8 centimeters) long with a wingspan of 7 inches (18 centimeters), the bats bite and then lap up blood from livestock or other animals at night.	D'environ 3 pouces (8 centimètres) de long avec une envergure de 7 pouces (18 centimètres), les chauves-souris mordent et lapent le sang du bétail ou d'autres animaux la nuit.	D'environ 3 pouces (8 centimètres) de long avec une envergure avec les ailes ouvertes de 7 pouces (18 centimètres), les chauves-souris mordent et lapent le sang du bétail ou d'autres animaux la nuit.

Résultats : expression mal corrigée (expliquée) par TH

Paires hétéromorphes TA-TH

Ratio PIT positif

Expression idiomatique

source	machine	humain
Expression idiomatique	Traduction littérale	Correction discutable

source_EN	machine_FR	humain_FR
There is a growing chorus of athlete complaints.	Il y a un chœur croissant de plaintes d'athlètes.	On assiste à un concert de voix grandissant de plaintes d'athlètes.

Résultats : hallucination non corrigée par TH

Distance euclidienne TH
3 fois la DE moyenne
caractères, mots et VIN

Modalité	machine	humain
Expression idiomatique	Traduction littérale abusive	Pas réellement corrigée

source_EN	machine_FR	humain_FR
About 15,000 members were expected to meet up in Red Deer on April 9 to cast a thumbs-up or thumbs-down on Kenney's leadership	Environ 15 000 membres devaient se réunir à Red Deer le 9 avril pour lever ou baisser le pouce sur le leadership de Kenney	Environ 15 000 membres devaient se réunir à Red Deer le 9 avril pour montrer le pouce en l'air ou vers le bas à l'égard du leadership de M. Kenney

Résultats : hallucination non corrigée par TH

Distance euclidienne TH
3 fois la DE moyenne
caractères, mots et VIN

Modalité	machine	humain
Expression idiomatique	Traduction littérale	Non corrigée

source_EN	machine_FR	humain_FR
says the decision " reeks of desperation " and that it's time to call an immediate leadership race	dit que la décision " pue le désespoir " et qu'il est temps d'appeler une course à la direction immédiate	dit que la décision « pue le désespoir » et qu'il est temps d'appeler une course à la direction immédiate

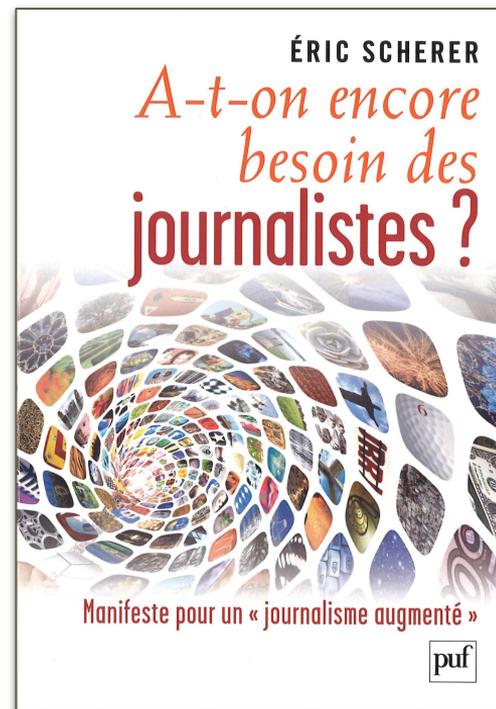
Discussion

Autant nos analyses par sacs de mots que par cinq paires de segments montrent intervention humaine importante.

L'humain demeure essentiel pour compléter la machine (traduction augmentée; journalisme augmenté).

Dès 2011, Éric Scherer plaidait pour un «journalisme augmenté».

Par contre, interventions non systématiques en révision. Signe que la cohabitation humain/machine n'est pas encore rodée, bien aménagée.



Conclusion: retour sur question de recherche

La TA est-elle appropriée pour répondre aux besoins de la Presse Canadienne?

On ne peut pas encore y répondre. Les risques d'erreurs sont par contre très présents en raison de la tendance à l'équivalence formelle (Martikainen et Kübler, 2016) qui produit chez les journalistes un nombre important de calques (détectés par Antidote).

On ne semble pas encore avoir réussi à combiner rapidité de la machine avec la qualité de l'humain.

Quel est le compromis qu'on est prêt à faire pour utiliser des machines?

Il faudrait approfondir cette étude.

Conclusion: pour aller plus loin

- Entrevues et observation des artisans de la PC.
- Analyse plus détaillée:
 - Typologie et criblage des améliorations à la TA, des erreurs de TA non corrigées en post-édition (traductions littérales abusives) et des améliorations obligatoires et facultatives de la TA en post-édition
 - Comparaisons entre TH intégrales (d'autres journalistes de la PC n'utilisant pas Ultrad) et post-éditions de TA
 - Automatiser la recherche de calques et d'interférences?
- Apport des traducteurs professionnels? Formation?
- Avenir de Ultrad : intégration dans CMS ou plugin? Utilisation de l'apprentissage automatique adapté à la PC?

Remerciements

Service français, Presse Canadienne:

- Geneviève Rossier, éditrice et directrice générale
- Marie-Ève Ouellette, gestionnaire du développement numérique