

Brest, 6 mars 2020


# Pour une approche raisonnée de la traduction automatique : *no more rage against the machine !*



Rudy Loock

STL Université de Lille

# ELLE EST LÀ !



2

- ▶ 2015 : arrivée de la **traduction automatique neuronale**
- ▶ **Progrès réels**, notamment fluidité de la langue cible
- ▶ **Transforme le marché** (méthodes de travail + modèle économique) : technologie **« disruptive »**
  - ▶ Pour la 1<sup>ère</sup> fois en 2018 : nombre d'entreprises de services linguistiques ayant recours à la TA en Europe > 50% (Tiny 2018)
  - ▶ Ne ralentit pas le marché des LSPs : +7% en 2018 (46,5 milliards \$)/+6,6% en 2019 (49,6 milliards \$)\* ; +28% emploi traduction/interprétation d'ici 2024\*\*

\*Source : Common Sense Advisory  
\*\* Source : US Bureau of Labor

# ELLE EST LÀ !

3

▶ Polarisation du débat :



- Ne veut pas en entendre parler
- Résultats toujours mauvais
- Pour les mauvais traducteurs
- Mépris pour la technologie

- Sur-évaluation de la qualité
- Équivalent à la traduction humaine
- Applicable à tous les domaines
- Traducteurs obsolètes
- Fascination pour la technologie

4

▶ Annonces marketing destinées au grand public

Microsoft reaches a historic milestone, using AI to match human performance in translating news from Chinese to English




Google Pixel Buds : les écouteurs-traducteurs qui vont venir à votre secours à l'étranger

Traducteurs et interprètes : tous chômeurs en 2025 ?

Achieving Human Parity on Automatic Chinese to English News Translation

5

▶ « Ratés » drôles (?) des outils de TA :

China's Tencent apologizes after WeChat translates Canadian flag into 'He's in prison'




Israel Eurovision winner Netta called 'a real cow' by Prime Minister in auto-translate fail

Facebook Apologizes for Translating Chinese Leader's Name to 'Mr. Sh\*thole'

6

▶ « Ratés » drôles (?) des outils de TA :


Elle a traduit « Hamlet » avec Google Traduction (et c'est rigolo)



“GOOGLE TRANSLATE SINGS”: POPULAR SONGS FIND NEW LYRICS

7

► « TA bashing » :




Chris Durban @ChrisDurbanFR - Nov 27, 2019  
Replying to @RudyLoock and @univ\_lille  
Rudy, deux Qc (1) qui sont les biotraducteurs qui interviendront lors de cette journée d'Etudes 7 Et (2) : parmi eux, y en a t il qui représentent autre chose que le vrac et le milieu de marché ?

Je suis consterné, aussi, de voir qu'une université va organiser sous peu une conférence sur le thème de la traduction automatique et des « biotraducteurs » (sic), comme si la norme était devenue la traduction automatique et qu'il fallait inventer une nouvelle appellation pour nous désigner, nous.

Liste de diffusion SFT, 2018

8

► « TA bashing » :



Sandrine de CUPPER - 25  
Traductrice professionnelle, 38 - FR  
Au vu des mêmes reproches de l'audio, j'ai du mal à croire que la traduction automatique fasse des heureux, chez les traducteurs professionnels du moins ->


André Thouvenel - 25  
"Biotraducteur" = traducteur professionnel qui utilise l'expression - Aut...  
Comment se dit le verbe anglais "traducteur traducteur" ?  
Avec, même avec la maîtrise des subtilités, traduire c'est trahir... La machine, la logique ou l'intelligence artificielle permettront à une autre...  
dépasser le stade de la phrase en un temps record et à l'ambulant...  
reconstitués dans un autre pays. Mais des outils de traduction...  
automatique n'ont pas cette humilité et cette sensibilité dont font preuve...  
le traducteur et l'interprète en tant que vecteurs d'une langue et d'une...  
raison et la portée des mots et des phrases qui les portent...  
Je ne parle même pas tout plus de la subtilité de la ponctuation dans les...  
textes et du jeu particulier dans l'usage des mots ou de certains éléments...  
irréalisables par une machine mais qui existent...  
Sans parler des enjeux épistémologiques de la machine en la matière et aussi...  
de celles et ceux qui veulent encore croire et pour longtemps en...  
l'humain.

Rudy Loock @RudyLoock  
#TO2020 @univ\_lille Sarah Van der Vorst | Le post-éditeur, un nouveau maillon fort du projet de traduction  
Olivier Manzoni @OlivierManzoni - Jan 31  
Replying to @RudyLoock and @univ\_lille  
Que signifie cette langue de bois appliquée à notre métier ? Sinon une nouvelle dégradation du travail des traducteurs pragmatiques ?

9

**Nécessités :**

- Approche raisonnée (Loock 2019)
- MT literacy (Bowker & Cro 2019)
- Demythification de la machine (Moorkens 2018)



**Et donc :**

- Maîtrise des différents enjeux (techniques, gestion de projet, déontologiques)
- Définition de la plus-value de la traduction humaine / biotraduction

10

**Maîtrise des enjeux techniques :**

- Enjeux :
  - Que sait faire la machine ?
  - Que ne sait-elle pas faire ?
  - ... avec des surprises !
- Quels sont les biais ?
  - Données dans les corpus
  - Algorithmiques
  - Sexistes et racistes

11

**Maîtrise des enjeux techniques :**

- Une boîte noire, certes
- Mais importance des données :
  - Outils « fit-for-purpose »
  - Importance des corpus :
    - Exemple : « Créations » lexicales :
      - Van Brussel et al. (2018) : traductions EN-DU : 15% des erreurs lexicales = mots qui n'existent pas
      - Macken et al. (2019) : impact négatif sur la compréhension dans 60% des cas

12

**Maîtrise des différents enjeux :**

- Biais algorithmiques
  - Risque de standardisation de la langue/perde de créativité avec sur-représentation des formes les + fréquentes (Vanmassenhove et al. 2019)
  - Reproduction des erreurs
  - Cf. mémoires de traduction

13

### Maîtrise des enjeux techniques :

#### ► Biais sexistes et racistes

- Difficultés pour les gestion des accords (Vanmassenhove 2018, Stanovsky et al. 2019)

Traduction anglaise (langue identifiée)	Traduction en français
The president has arrived. The nurse has arrived. The president asked the secretary to help her in the procedure.	Le président est arrivé. L'infirmière est arrivée. Le président a demandé à la secrétaire de l'aider dans la procédure.

- Reproduction des biais racistes

14

### Maîtrise des enjeux techniques :

#### ► « Racisme technique »



The Jews are tainting the blood of the Germans.  
Les Juifs souillent le sang Allemand.

Hitler était beaucoup plus gentil envers les Juifs qu'ils le méritaient.

Pour être honnête, le plupart des animaux n'ont [aucun péché](#) [à se reprocher](#).

birman > anglais :  
Xi Jinping > Mr. Shithole

Source : Reverso 2018

15

### Maîtrise des enjeux techniques :

#### ► Objectifs :

- être capable d'analyser/de comprendre les sorties de TA au-delà de la boîte noire
- se méfier des biais (et les corriger)
- déconstruire le discours marketing

- En route vers la **post-édition** !

16

### Maîtrise des enjeux liés à la gestion de projets :

#### ► Quand utiliser la TA ?

- Comment calculer les tarifs / la marge ?
- Un rôle élargi pour le post-éditeur (Van der Vorst 2020)

17

### Maîtrise des enjeux déontologiques :

- Quand/comment informer le client ?
- Confidentialité des données
- Empreinte écologique (Strubell et al. 2019)
- Corpus < traductions humaines : quelle reconnaissance du travail fourni ? (Kenny 2012)

18

### Définition de la plus-value de la biotraduction

#### ► Quelles différences entre traduction humaine et traduction automatique ?

- Nombreuses expériences comparent TH <> TA <> TAPE
- Machine Translationese / Post-edited (Lapshinova-Koltunski (2015), Thorat 2019, Daems et al. 2017, Vanmassenhove et al. 2019, Martikainen 2019, Loock 2018, 2020, De Clercq et al. 2020)

De Clercq et al. (2020)

19

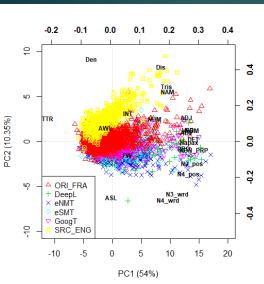


De Clercq et al. (2020)

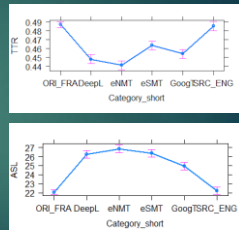
	GB_EN	FR
Business&Finance	6 136	36 964
Crime	43 710	93 347
Culture	46 839	78 897
Environment	32 367	88 574
Health	28 170	65 024
International News	29 168	65 354
Politics	46 901	98 540
Science & Technologies	47 213	97 252
Sports	43 766	97 367
Travel	50 056	97 351
TOTAL #tokens	374 326	818 670
#files	490	1 094



TA en FR



21



## Définition de la plus-value de la biotraduction

22

- Analyse du machine-translationese > informations pour la post-édition
- Vanmassenhove et al. (2019) : moindre richesse lexicale
- Loock (2020) : non traduction des adverbes (3,5% pour DeepL/10% pour eTranslation)

23

## Quelles conséquences pour l'enseignement de la traduction ?

24

- Défi majeur : enseignement de **la TA** (outil) mais aussi enseignement de **la traduction avec TA**
- Expériences nombreuses : Rossi (2017), Massey & Ehrensberger-Dow (2017), Moorikens (2018), Faria Pires (2018), Marikainen (2019), Yamadu (2019)
- Pose également la question de l'évaluation : qu'évalue-t-on ?

25

## Face à la TAN, plusieurs leviers (Loock 2019) :

1. Enseigner la traduction comme un **service commercial** que l'on rend à des clients
2. **Connaissance** et **maîtrise** des outils de TA pour une **approche raisonnée**
3. Dégager la **plus-value de l'humain** en encourageant la **créativité**

Objectif = « future-proofing the profession »

26

## 1. Enseigner la traduction comme un **service commercial**



Cf. Référentiel de compétences du réseau EMT

27

## 1. Enseigner la traduction comme un **service commercial**

Au-delà du transfert linguistique  
Gestion de projets  
Services complémentaires  
Développement des *soft skills*

=> Proposer des **projets de traduction à gérer**

28

**Consignes spécifiques** : terminologie spécifique, guides de style, localisation d'URL, référencement (SEO), format....

**Adaptation** :

Institute of Education  
LibGuides

**Zotero**

**UCL**

You may want to purchase extra space so that you can access the full-text online on any computer with internet access. Alternatively, you may want to save your files on the machine which is the network drive on the computers at UCL.

If you would like Zotero to find the resource in Explore, UCL's catalogue, add the URL, exactly as it appears here in the **Repository field**: [https://www.primo.hosted.axillabgroup.com/openurl/UCL/UCL\\_VUZ?](https://www.primo.hosted.axillabgroup.com/openurl/UCL/UCL_VUZ?)

→ Université de Lille

29

## 1. Enseigner la traduction comme un **service commercial**

Importance de la **mise en situation** :

- Au-delà du transfert linguistique (projets)
- Autoriser le travail avec tous les outils à disposition (y compris la TA pendant les examens !)
- Mise en place de dispositifs spécifiques de mise en situation : **agence virtuelle de traduction, Skills Lab, approche par projets, Junior Entreprise**

**Exemple** : TSM Skills Lab, Université de Lille

Agence virtuelle de traduction au sein de la formation :

1. Candidature sur différents postes : gestion de projets, préparation fichiers/ressources, traduction, révision, communication (CV, lettre de motivation)
2. Entretien de recrutement
3. Semaine intensive : envoi de projets de traduction multiples avec consignes le lundi matin ; livraison vendredi soir avec facturation (fictive)





### Exemple : TSM Skills Lab, Université de Lille

- Mobilisation de l'ensemble de leurs compétences au service de projets communs
- Prise de décision (notamment pour la TA)
- Travail collaboratif => développement des *soft skills*  
Relation client, organisation, gestion des imprévus
- Retour sur expérience : « métacognition »



32

## 2. Connaissance & Maîtrise des outils de TA

- Découvrir la TA comme outil d'**aide** à la traduction dès que possible
- Formation à la pré-/post-édition
- Comprendre ce que la machine (ne) sait (pas) faire afin de dégager sa **plus-value** :
  - **Expérimentations** à l'aide d'outils génériques et personnalisés (comparaisons TA vs. pas TA)
  - Analyse des **« erreurs »** des outils et découverte des **biais** (sexistes, racistes, algorithmiques)

33

## 2. Connaissance & Maîtrise des outils de TA

- Intégration dans la **gestion de projets**
- Questions de **déontologie**
- Laisser les étudiants **libres** de l'utiliser, y compris lors d'examens !
- Lecture de la **littérature scientifique** sur le sujet
- **Conférences** sur le sujet avec regards croisés
- Mais en revanche, **cesser de faire rire** avec la TA !

34

## 3. Outils permettant de développer la créativité

- Un des risques identifiés de la post-édition : standardisation de la langue, perte du « style »
- Cf. **machine translationese, post-editedese** (Thorat 2019, Daems et al. 2017, De Clercq et al. 2019)
- **Sensibilisation** à l'**usage** de la langue, au-delà du grammaticalement correct !

35

## 3. Outils permettant de développer la créativité

- Développer le **rédactionnel**, l'**idomaticité**, la **variété lexicale**
- Viser l'**homogénéisation linguistique** avec la langue originale

36

## 3. Outils permettant de développer la créativité

Quoi de mieux pour cela que les **outils de corpus** !

- Mise au jour de différences inter-langagières LS/LC originales
- Mise au jour de différences intra-langagières langue originale/langue traduite

< approche **« traductologie de corpus »** : Loock (2016, 2018, 2019)

## Exemple : l'ordre des mots (Loock 2018)

37

- (1) The president has bought a new computer./La présidente a acheté un nouvel ordinateur.
- (2) It is a new computer that the president has bought. / C'est un nouvel ordinateur que la présidente a acheté.
- (3) What the president has bought is a new computer./Ce que la présidente a acheté, c'est un nouvel ordinateur.
- (4) a. The president, she, has bought a new computer./La présidente, elle, a acheté un nouvel ordinateur.  
b. She, has bought a new computer, the president./Elle, a acheté un nouvel ordinateur, la présidente.
- (5) It is obvious that the president has bought a new computer./Il est évident que la présidente a acheté un nouvel ordinateur.

NON-STANDARD

SVO/STANDARD

STRUCTURE CLIVÉE

STRUCTURE PSEUDO-CLIVÉE

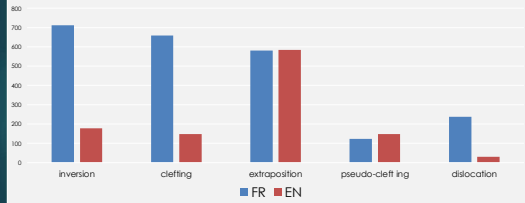
DÉLOCATION (DROITE/GAUCHE)

EXTRAPOSITION



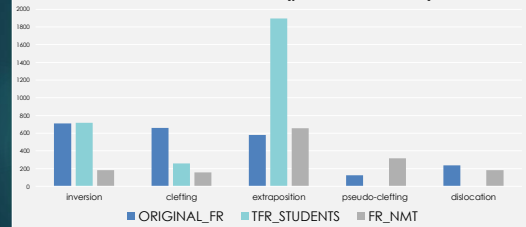
39

Ordre des mots non SVO (par million de mots)



40

Ordre des mots non SVO (par million de mots)



## Extraposition

Il est possible que le prochain ordinateur de votre enfant fonctionne avec un logiciel Microsoft, mais ce n'est peut-être pas le Windows auquel vous êtes habitué.

< Your kid's next computer may say it's powered by Microsoft's software, but it may not be the Windows you're used to.

Il vous sera également impossible d'installer vos propres applications.

< You won't be able to install your own apps.

Il est peu probable que l'on clique sur des sites qui n'apparaissent pas dans les premiers résultats.

< Sites that do not show up among the first hits are unlikely to be clicked on.

42

## ▶ Très bon outil de sensibilisation :

- ▶ Homogénéisation linguistique langue originale/ langue traduite (invisibilité)
- ▶ Différence avec la machine : que contrôler durant la post-edition/révision ?
- ▶ Exercices d'application : introduction obligatoire d'ordre des mots non standard

## CONCLUSIONS

43

- ▶ Puisque la traduction automatique est là, il faut apprendre à travailler avec
- ▶ Il faut y former les traducteurs et les interprètes de demain (dès le 1er jour !)
- ▶ La plus-value de l'humain par rapport à la machine est réelle : il faut la définir, l'exploiter, et y sensibiliser les étudiants

**Approche scientifique, raisonnée, dépassionnée**

## CONCLUSIONS

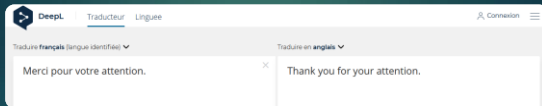
44

*"The most effective translators, at least for complex bodies of text, are neither human nor machine. They are, in fact, centaurs — humans augmented with the support of technology."*



(Simone Stolzoff, technology reporter at Quartz, Dec. 2018)

45



### Contact :

rudy.loock@univ-lille.fr

<https://pro.univ-lille.fr/rudy-loock/>

[@RudyLoock](https://twitter.com/RudyLoock)