

#AFFUMT2021, 8 April 2021

*Translationese, Machine Translationese,
Post-editeese, et invisibilité des futur(e)s
traducteurs/trices*

*Translationese, Machine Translationese,
Post-editeese, and the invisibility of
translators-to-be*

Rudy Look



Bruno Catalano



Introduction

- Specialized translation: market expectations = **translator invisibility**



- Holy Grail : **linguistic homogenization** with original language



Introduction

How?

- Focus on writing skills in the target language
- Use of technological translation tools to get inspiration from experts
 - Translation memories (< translation experts)
 - Specialised electronic corpora (< domain experts)

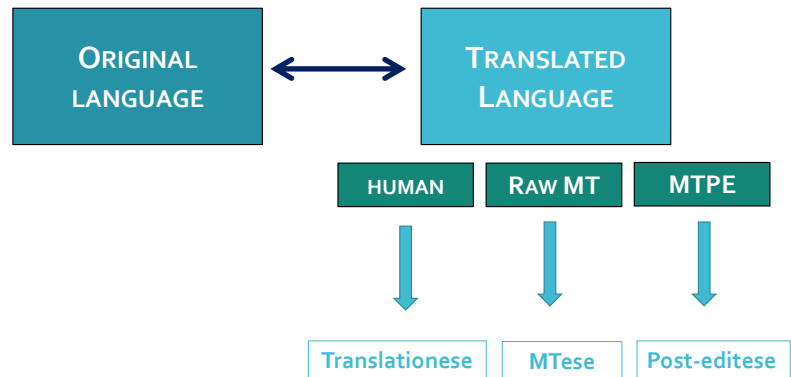
Introduction

How?

- Challenge: how to make translation students aware of the **gap** between original language and their productions (translation + MTPE)?
 - ⇒ Identification of linguistic differences
 - ⇒ Sensitization
 - ⇒ Remediation during revision, PE

Methodology

- Analysis of translated texts (human/machine) as **electronic corpora**
- Identification of **intra-language differences**

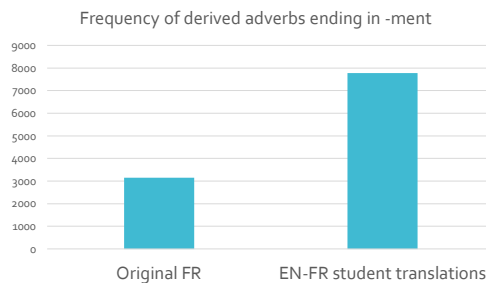


Translationese

- < corpus-based translation studies
- Baker (1993)
- Comparison original vs. translated texts
- Differences almost systematic:
 - Fewer contractions in EN translated texts (Baker & Olohan 2000)
 - Lower lexical variety and density in EN and ZH translated texts (Laviosa 2002, Xiao 2010)
 - Fewer borrowings in EN-IT translated pods (Bernardini & Ferraresi 2011)
 - Fewer PP extrapositions in DE/NL translated texts (De Sutter & Van de Velde 2011)
 - More existential constructions, derived adverbs in EN-FR translated texts (Cappelle & Looock 2013, Looock 2016)
 - ...

Translationese

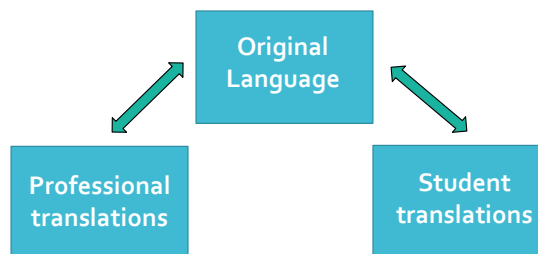
- Interpretations: third code/variation (« translation features ») vs. translationese?
- Raging debate among researchers
- But nice sensitization tool for students
- Possible comparisons with professional translations



Translationese:

De Sutter et al. 2017

• Data

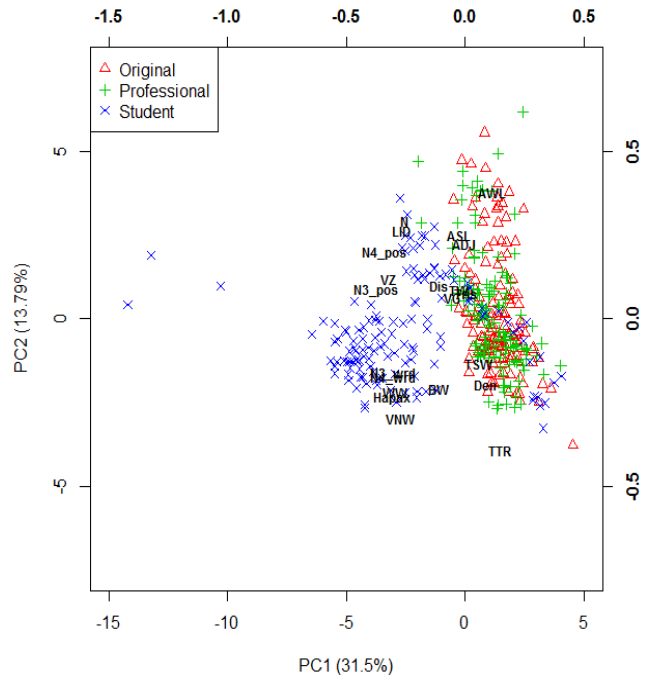


- EN-FR fiction texts (master's level), FR-NL press texts (undergraduate levels)
- **30 metrics:** TTR, lexical density, average word/sentence length, hapax/dis/tris legomena, n-grams, PoS tags...
- 1 text = 1 linguistic profile
- Data visualization with principal Component Analysis

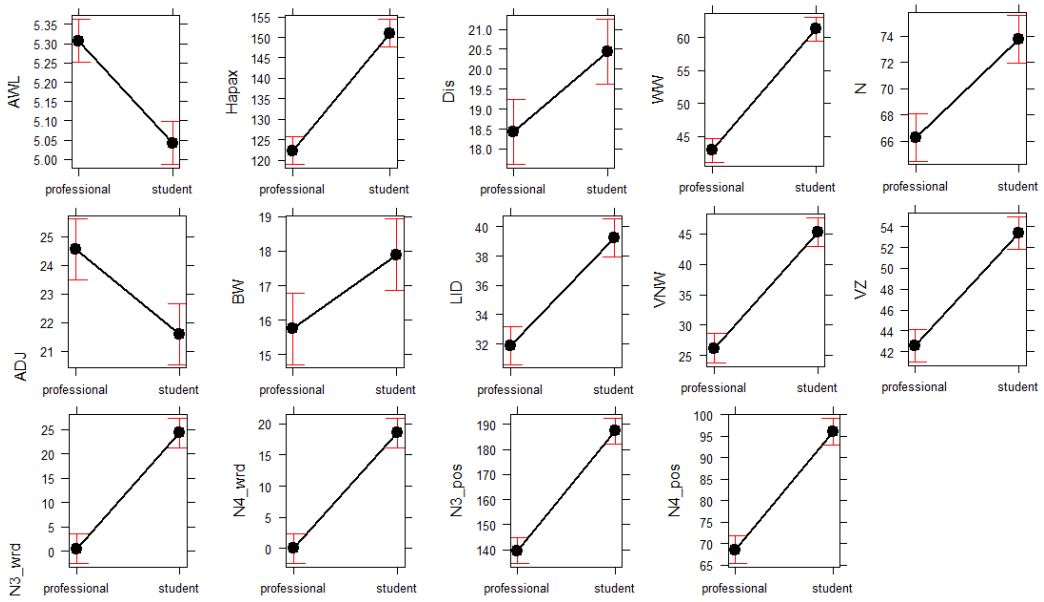
FR-NL press texts

Translationese:

De Sutter et al. 2017



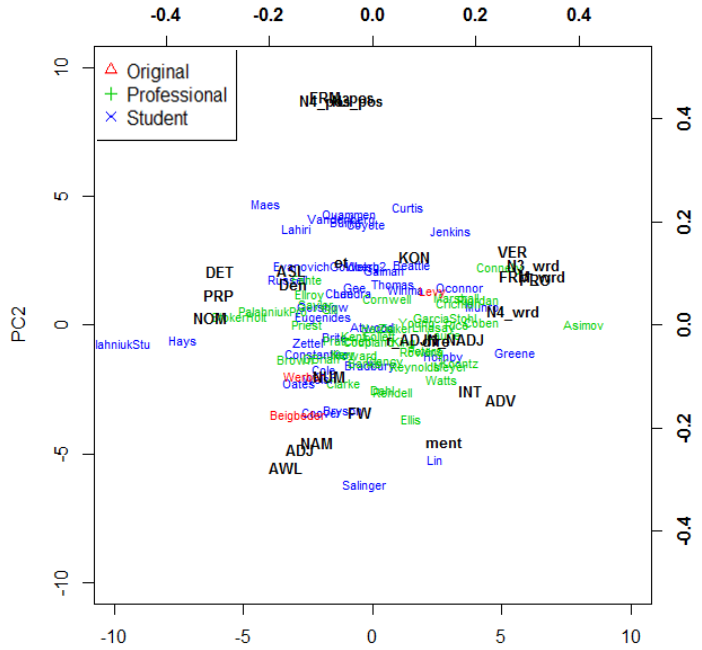
FR-NL press texts



EN-FR fiction texts

Translationese:

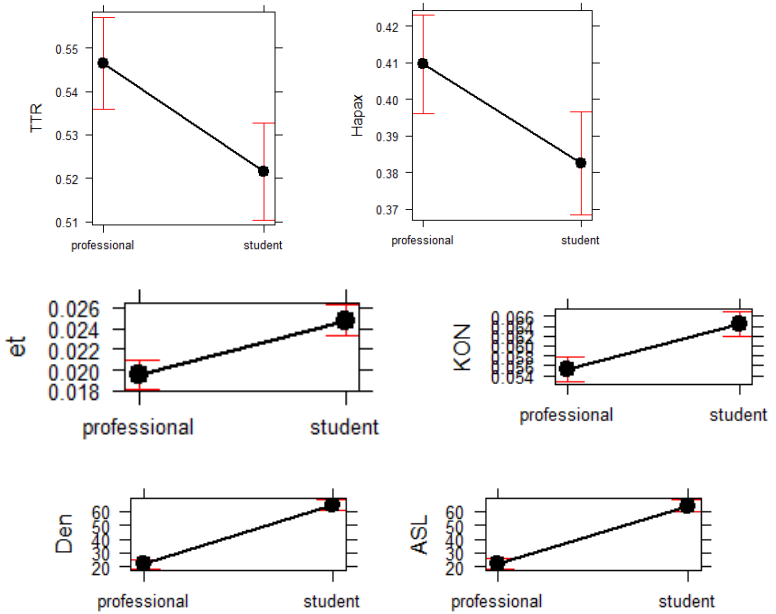
De Sutter et al. 2017



EN-FR fiction texts

Translationese:

De Sutter et al. 2017



EN-FR fiction texts

Translationese:

De Sutter et al. 2017

Conclusion:

- Students can visualize linguistic differences with professional translations and original language
- ≠ automatic correction of students' translation tasks
- But objective, data-based feedback on what needs to be improved



Machine- Translationese/ Post editese

- NMT is here to stay!
- Real progress, in particular TL fluency
- Need to fight against generalizations, false ideas and perceptions
- Need for debunking, demythification
- Professionals and students need to define their added value over the machine

= MT literacy



Machine- Translationese/ Post editese

• MT output evaluation:

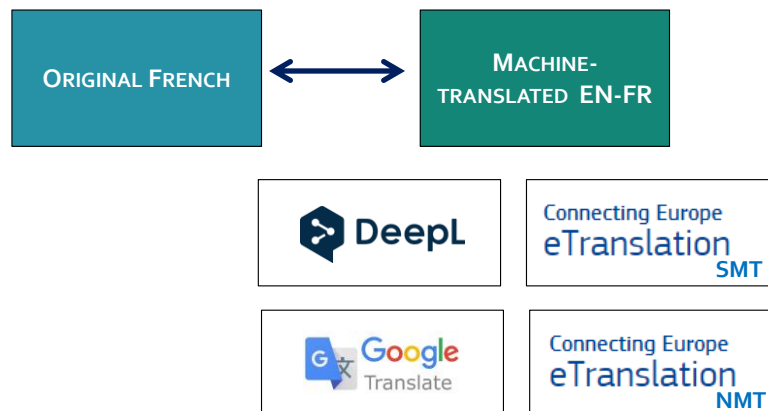
- **Metrics** (BLEU, METEOR, ROUGE...)
- **Human evaluation:**
 - ranking by professionals or non-professionals
 - amount of necessary post-editing (PE)
 - error identification and classification
 - linguistic evaluation
 - Series of sentences
 - Series of texts

Corpus-based
approach

Machine- Translationese

Loock 2018, 2020 / De
Clercq et al. 2020,
2021

- Example: experiment on 4 MT systems for EN-FR translations



Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021

- Example: experiment on 4 MT systems for EN-FR translations
 - **Langue-dependent features:** frequency of *chose, et, avec, dire* (all inflections), existential constructions, -ment adverbs
 - **Language-independent features:** TTR, lexical density, POS-tags, hapax/dis/tris legomena, n-grams, average word/sentence length

Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021

Data:

TSM press corpus

	Ori US_EN	Ori GB_EN	Ori FR
Business&Finance	27 487	6 136	36 964
Crime	44 315	43 710	93 347
Culture	30 570	46 839	78 897
Environment	41 500	32 367	88 574
Health	34 790	28 170	65 024
International News	33 767	29 168	65 354
Politics	45 840	46 901	98 540
Science & Technologies	45 269	47 213	97 252
Sports	45 156	43 766	97 367
Travel	40 748	50 056	97 351
TOTAL #tokens	389 442	374 326	818 670
#files	437	490	1 094
TOTAL TSM press corpus			1 582 438

Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021

Data:

Business&Finance	
Crime	
Culture	
Environment	
Health	
International News	
Politics	
Science & Technologies	
Sports	
Travel	
TOTAL #tokens	
#files	
TOTAL TSM press corpus	

TSM press corpus

Ori US_EN	Ori GB_EN	Ori FR
27 487	6 136	36 964
44 315	43 710	93 347
30 570	46 839	78 897
41 500	32 367	88 574
34 790	28 170	65 024
33 767	29 168	65 354
45 840	46 901	98 540
45 269	47 213	97 252
45 156	43 766	97 367
40 748	50 056	97 351
389 442	374 326	818 670
437	490	1 094
		1 582 438

Translation into FR

Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021

Data:

Business&Finance	
Crime	
Culture	
Environment	
Health	
International News	
Politics	
Science & Technologies	
Sports	
Travel	
TOTAL #tokens	
#files	
TOTAL TSM press corpus	

TSM press corpus

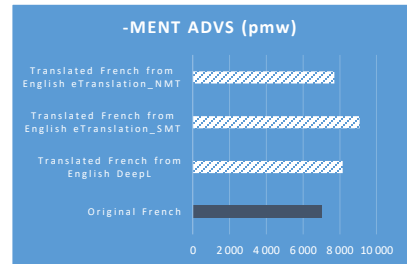
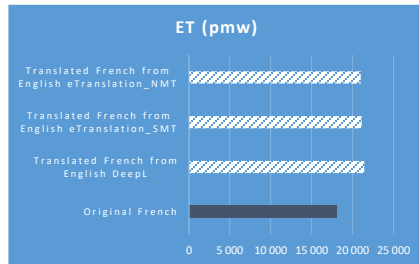
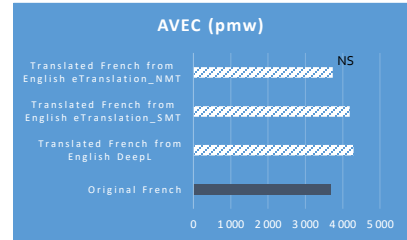
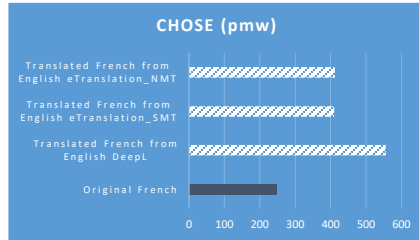
Ori US_EN	Ori GB_EN	Ori FR
27 487	6 136	36 964
44 315	43 710	93 347
30 570	46 839	78 897
41 500	32 367	88 574
34 790	28 170	65 024
33 767	29 168	65 354
45 840	46 901	98 540
45 269	47 213	97 252
45 156	43 766	97 367
40 748	50 056	97 351
389 442	374 326	818 670
437	490	1 094
		1 582 438

Translation into FR

• Results > Language dependent features

Machine-Translationese

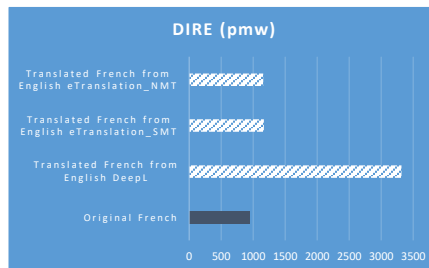
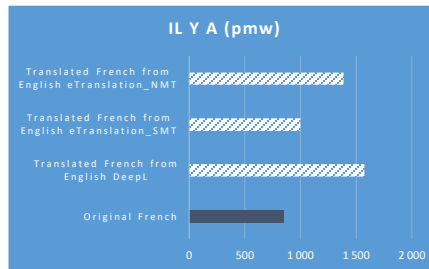
Look 2018, 2020 / De Clercq et al. 2020, 2021



• Results > Language dependent features

Machine-Translationese

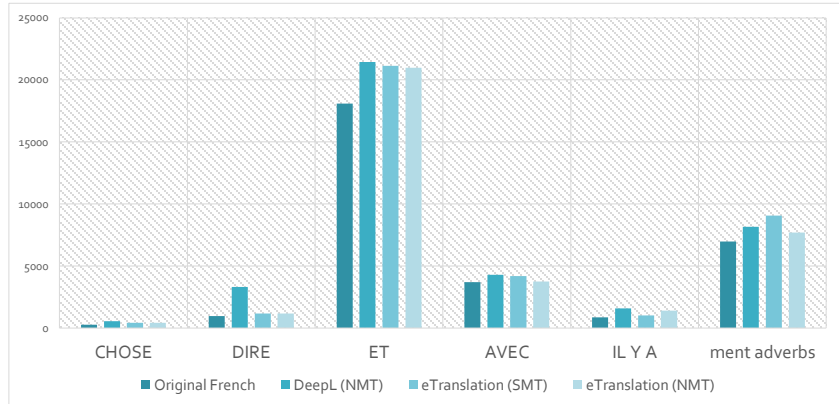
Look 2018, 2020 / De Clercq et al. 2020, 2021



- Results > Language dependent features

Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021



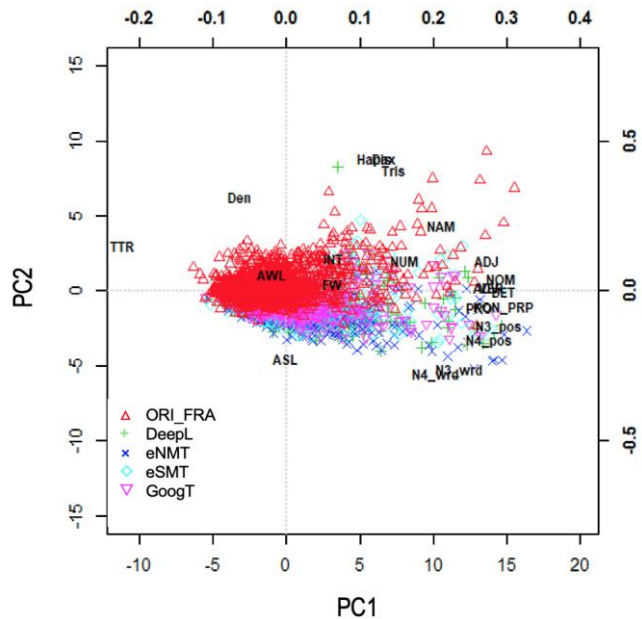
Note:

- Source language interference not the sole explanation
- Qualitative analysis shows situation is complex
e.g.: -ly adverbs not translated in 3.5% of cases (DeepL), 10% (eTranslation)

- Results > Language independent features

Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021

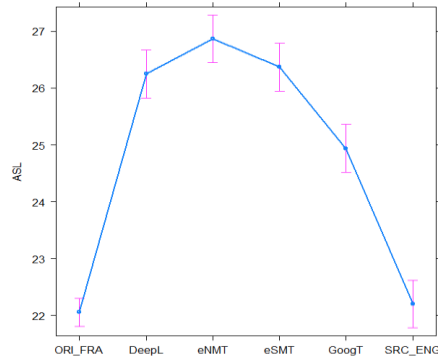


- Results > Language independent features

Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021

Average Sentence Length:



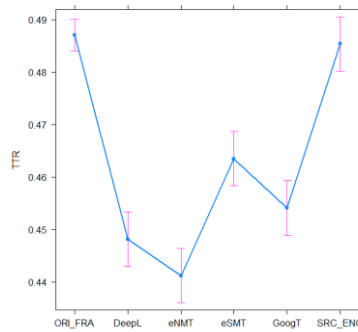
Text length:

+18.2% (DeepL)
 +15.2% (GoogT)
 +19.1% (eSMT)
 +20.6% (eNMT)

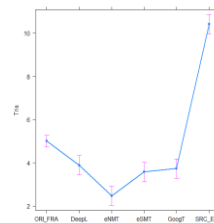
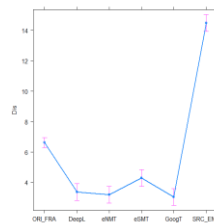
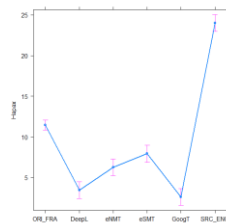
- Results > Language independent features

Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021



Still, MT can be « creative »:
cavalage, vampiriquement, bienbouffeur, vortir, tonnamment, inquieté...

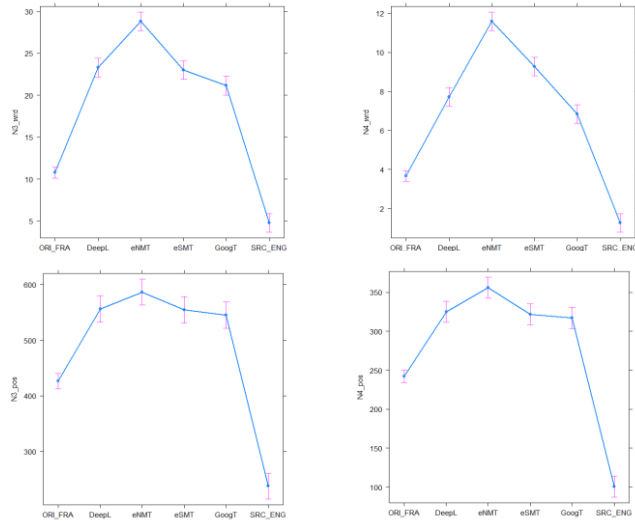


- Results > Language independent features

Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021

N-gram features (3- and 4 grams)



Machine-Translationese

Loock 2018, 2020 / De Clercq et al. 2020, 2021

- Machine-translationese is a reality
- Potential quality issue
- Information on what to check during PE

= MT literacy, empowerment, principled approach

Post-editeuse?

Daems et al. 2017

- EN-NL translations
 - Readers fail to distinguish HT and MTPE
 - Investigation of 55 linguistic features also fails to distinguish HT and MTPE

Conclusion of the study:
post-editeuse = translationese?

Post-editeuse?

Toral 2019

- EN/ES-DE, EN-FR, ZH-EN, ES-DE translations
 - TTR (lexical variety) : HT > PE > MT
 - Lexical density: HT > PE = MT
 - Sentence length: PE more similar to source texts

Conclusion of the study:
post-editeuse = « exarcebated translationese »

CONCLUSIONS

- Measuring the gap between:

ORIGINAL LANGUAGE

HUMAN TRANSLATION

MACHINE TRANSLATION
(RAW/MTPE)

can help sensitize students to linguistic differences that can be 'corrected' during revision/PE:

- Translationese
- Machine translationese
- Post-editedese

= corpora as sensitization tools

CONCLUSIONS

- Check-list of linguistic features to work on with students:

e.g. EN-FR human translations:

- lexical variety/density
- frequency of coordinator *et* and conjunctions
- sentence length

e.g. EN-FR MT

- lexical variety
- linguistic features showing cross-linguistic differences: existential constructions, derived adverbs, *dire...*
- made-up words

Contact :



rudy.loock@univ-lille.fr



@RudyLoock

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

attribué par un paiement de carte bancaire sur notre propre site **merci pour votre attention** ->-> Actuellement le service de micro paiement Media Pay est
 Sylvie et Michel sont des hôtes formidables et dévoués. ->-> **Merci pour votre attention** pendant ma période de blocage. ->-> Cette semaine, bien que
 juridique + traductrice ->-> Références disponibles. ->-> **Merci pour votre attention** ->-> Natalie ->-> Jeune femme sérieuse et dynamique ne
 mon souhait et je commence à le tester avec mon chien. ->-> **Merci pour votre attention** très agréable avec votre caractère et 2 confiseries à l'inté
 jement tous personnes qui s'engage sur l'allée indvise ? ->-> **merci pour votre attention** à mon souci, mon voisin n'injurie pas ceux qui passent sur l'allée et
 en Angleterre je peux assurer un bon niveau en Anglais. ->-> **Merci pour votre attention** . Po - disponible à partir du 16 octobre Marie ->-> Jeune diplo
 tments Contexte et réglementations Bourgel-du-Lac, 07/12/2011 **Merci pour votre attention** CETE de Lyon Ministère de l'écologie, du Développement durable
 tout le monde et en particulier par les esprits les plus étroits *** **Merci pour votre attention** ->-> A bord, le temps d'une mini croisière. ->-> Choisir C
 En attente d'une réponse pour une formation de gendarme G.A.V **Merci pour votre attention** . En attente d'une réponse je vous pris d'agréer Madame, Monsieur,
 en janvier 1973 pour n'avoir que 4.30 € 40 ans plus tard. ->-> **Merci pour votre attention** ->-> J'ai un livret c e pt en date de 1963 3642 franc pourriez v
 he (Lh 5, B) ->-> A vous tous et à vous toutes, bienvenue, et **merci pour votre attention** bienveillante. ->-> +Claude Rault. ->-> Evêque de Laghou
 s) est ce que je fais maintenant le charge en phase liquide ou gaz **merci pour votre attention** ->-> Certains des matériaux cités dans cette rubrique servent
 me 1 ou du moins s'en retirer, et aussi revenir du Franc 1 ->-> **Merci pour votre attention** . Meilleures salutations, do http://ma68.org ->-> Post-scriptur
 e dessins et de peintures en suivant les liens ci dessous. ->-> **Merci pour votre attention** et à vos crayons 1 ->-> PREPARATION CONCOURS ECOLES
 les moi ou cela se trouve et je rappellerai directement. ->-> **Merci pour votre attention** Maxime Davault ->-> Annonces immobilières sur la Côte d'Az
 gnages et conseils sur le guide du savoir Web Libre.org ->-> **Merci pour votre attention** le téléphone me demande le code SIM chaque fois que je met ma
 ent publier mon récit des événements sur votre site web. ->-> **Merci pour votre attention** et votre partage. ->-> ->-> Ce tableau d'affichage en ligne est dispo
 voyez plus que le smaj j'ai été déclarée au réel. est bien normal **merci pour votre attention** ->-> Le montant de votre retraite dépend de votre âge mais au
 par l'Agar Agar, ou du moins, une partie, ce serait un + ->-> **Merci pour votre attention** , vos futures réponses et vos opinions. ->-> J'ai franchement h



To download the references mentioned in the presentation: