



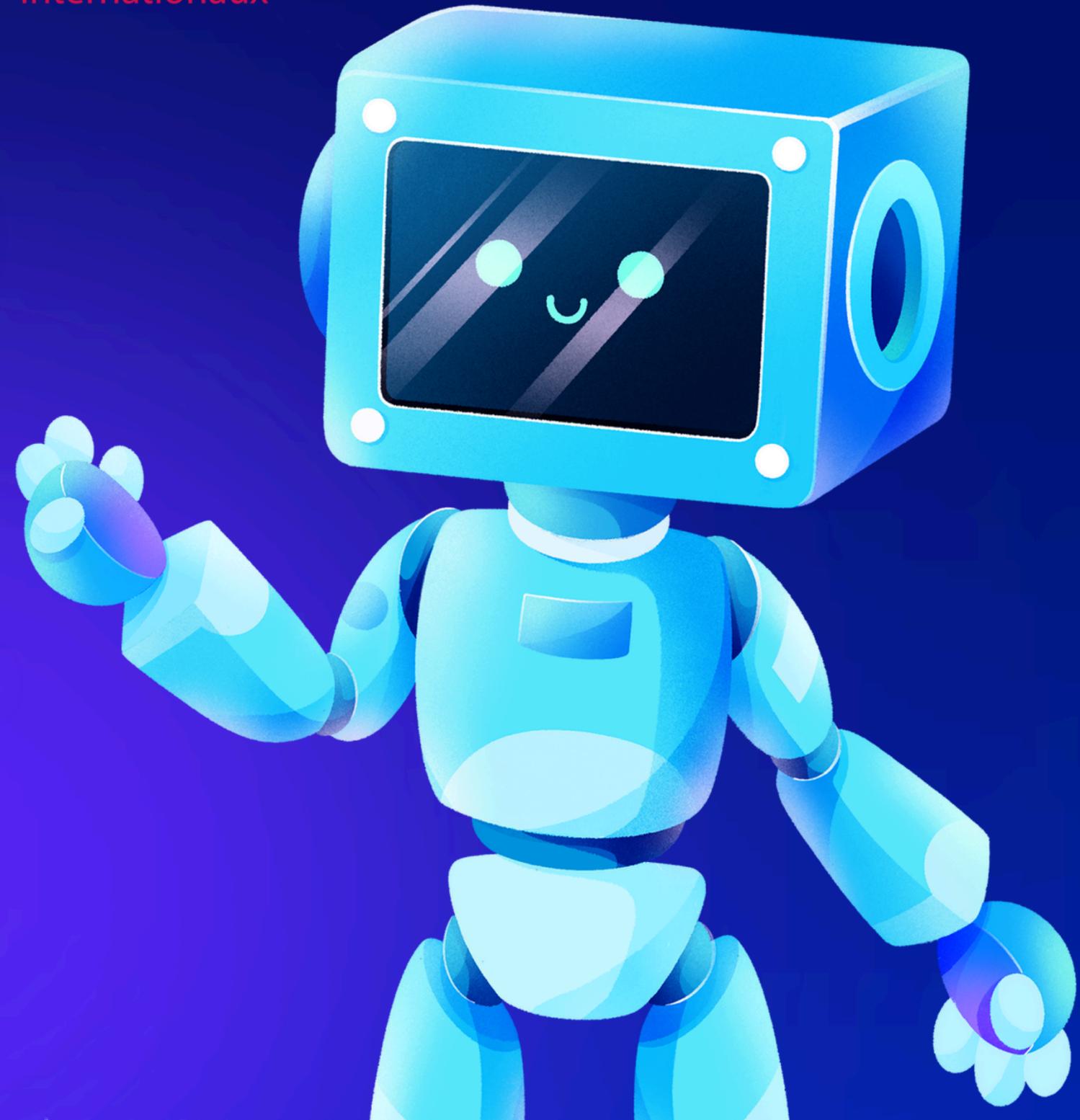
Faculté
de Traduction
et d'Interprétation

Ecole d'Interprètes
Internationaux

POSEDION : UN OUTIL POLYVALENT POUR L'ÉTUDE DU PROCESSUS DE POST-ÉDITION

Loïc de Faria Pires
Service ELLIT

Outils et méthodes de recherche, centre TELL, 13/11/2024





POURQUOI UN TEL OUTIL ?

- Analyse du **processus** de PE (>< produit)
 - Vise majoritairement à évaluer l'effort de PE
 - Traditionnellement divisé en 3 composantes (Krings & Koby, 2001)
 - Effort temporel
 - Effort technique
 - Effort cognitif
- 



TROIS DIMENSIONS DE L'EFFORT

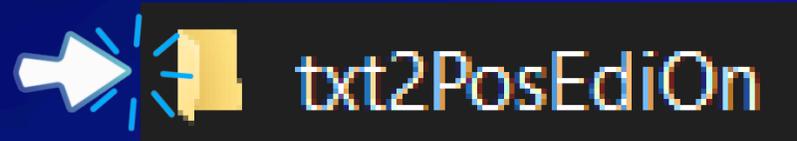
- **Effort temporel : mesure du temps de PE**
 - Pour un texte donné, quelle est la durée de PE chez les post-éditeurs en fonction du moteur utilisé ?
 - Quelle TA demande l'intervention humaine la moins chronophage ?
 - **Effort technique : mesure des opérations "manuelles"**
 - Pour un texte donné, avec une TA donnée, quelle est la quantité d'insertions, de suppressions, de frappes clavier...?
 - Quelle TA permet le nombre d'opérations techniques le plus bas ?
 - **Effort cognitif : mesure de la réflexion nécessaire** (le + ardu à mesurer)
 - P. ex., pour un texte donné et une TA données, quelle est la quantité et la longueur des pauses lors du processus ?
 - Aussi *think-aloud protocols*, etc. (pas traités par le logiciel)
- 

POSEDION

- Jusque récemment, outils pas très efficaces / compliqués à mettre en œuvre dans notre environnement informatique (ex. Translog)
- Outil présenté par Oliver *et al.* (2020)
 - Script Python avec paramétrage personnalisé
 - Exécutable via un simple .exe
- Mesures proposées
 - Temps de PE
 - Frappes clavier
 - Pausés : nombre et longueur
 - Métriques automatiques permettant d'estimer la qualité de la TA brute : BLEU, HTER, Comet...



COMPOSANTES : 1-CONVERTIR LES TEXTES



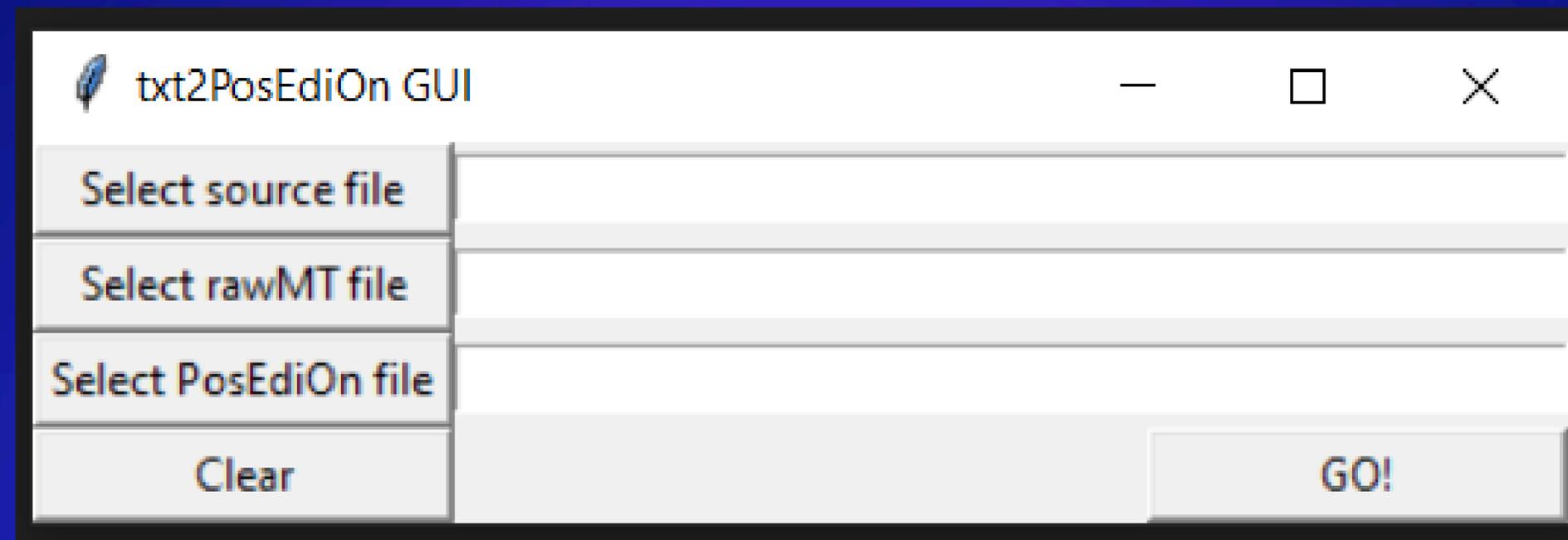
Pré-requis : disposer d'un .txt contenant le TS, et un contenant la TA brute, avec un retour à la ligne par segment

```
texte source.txt - Bloc-notes
Fichier Edition Format Affichage Aide
you: hello
receptionist: hello, hello. be a dear and fill out this form.
you: um, i don't remember how i got here...
receptionist: i'm sorry to hear that. i need you to sign on this dotted line here.
the receptionist hands you a pen and a clipboard.
"purgatory is spelled wrong"
you: um, purgatory is spelled wrong here

gpt.txt - Bloc-notes
Fichier Edition Format Affichage Aide
toi : bonjour
réceptionniste : bonjour, bonjour. Sois cher et remplis ce formulaire.
toi : euh, je ne me souviens pas comment je suis arrivé ici...
réceptionniste : je suis désolée d'entendre ça. J'ai besoin que tu signes ici sur cette ligr
la réceptionniste te tend un stylo et un carnet.
"le purgatoire est mal orthographié"
toi : euh, le purgatoire est mal orthographié ici.
```

COMPOSANTES : 1-CONVERTIR LES TEXTES

Ensuite, txt2posedion nous permet d'obtenir un fichier exploitable (format .sqlite) qui contient les textes et dont on choisit l'emplacement (!)



COMPOSANTES : 2-CRÉER LE PACK



config.yaml	01-10-24 15:27	Fichier source Yaml
PosEdiOn.exe	09-09-21 19:09	Application
PosEdiOn.py	09-09-21 19:08	Fichier source Pyth...
projet-3.sqlite	01-10-24 15:25	Fichier SQLITE

- 1) Déplacer le .sqlite fraîchement créé dans ce dossier
- 2) Double-cliquer sur config.yaml pour paramétrer

```
1 Project:
2   file: projet-3.sqlite
3
4 Languages:
5   source: eng
6   target: fra
7
8 Size:
9   height: 10
10  width: 80
11
12 Behaviour:
13   allowEditSL: True
14
15 Font:
16   font: courier 12
17
18 Chronometer:
19   status: show
20   #possible values: show / hide
```

Utiliser le même nom que votre docu .sqlite (liaison entre fichiers)

Indiquer langue source / cible

Paramétrage de la taille de la fenêtre pour les post-éditeurs

Choix de la police

Afficher / masquer le chrono pour les participants

COMPOSANTES : 3-ENVOYER LE PACK AUX PARTICIPANTS

Créer un dossier compressé contenant les 4 éléments une fois le .yaml sauvegardé

L'envoyer aux participants (email, Moodle, ...)

Eux le dézippent, double-cliquent sur PosEdiOn et le programme se lance

config.yaml	01-10-24 15:27	Fichier source Yaml
PosEdiOn.exe	09-09-21 19:09	Application
PosEdiOn.py	09-09-21 19:08	Fichier source Pyth...
projet-3.sqlite	01-10-24 15:25	Fichier SQLITE



```
PosEdiOn
you: hello

toi : bonjour

0:00:03 1/89 PAUSE << ACCEPT >> GO TO Clear
```



COMPOSANTES : 4-ANALYSE DES DONNÉES

Les post-éditeurs valident les segments ("Accept") une fois ceux-ci post-édités, et ferment le programme.



 PosEdiOn-analyzer

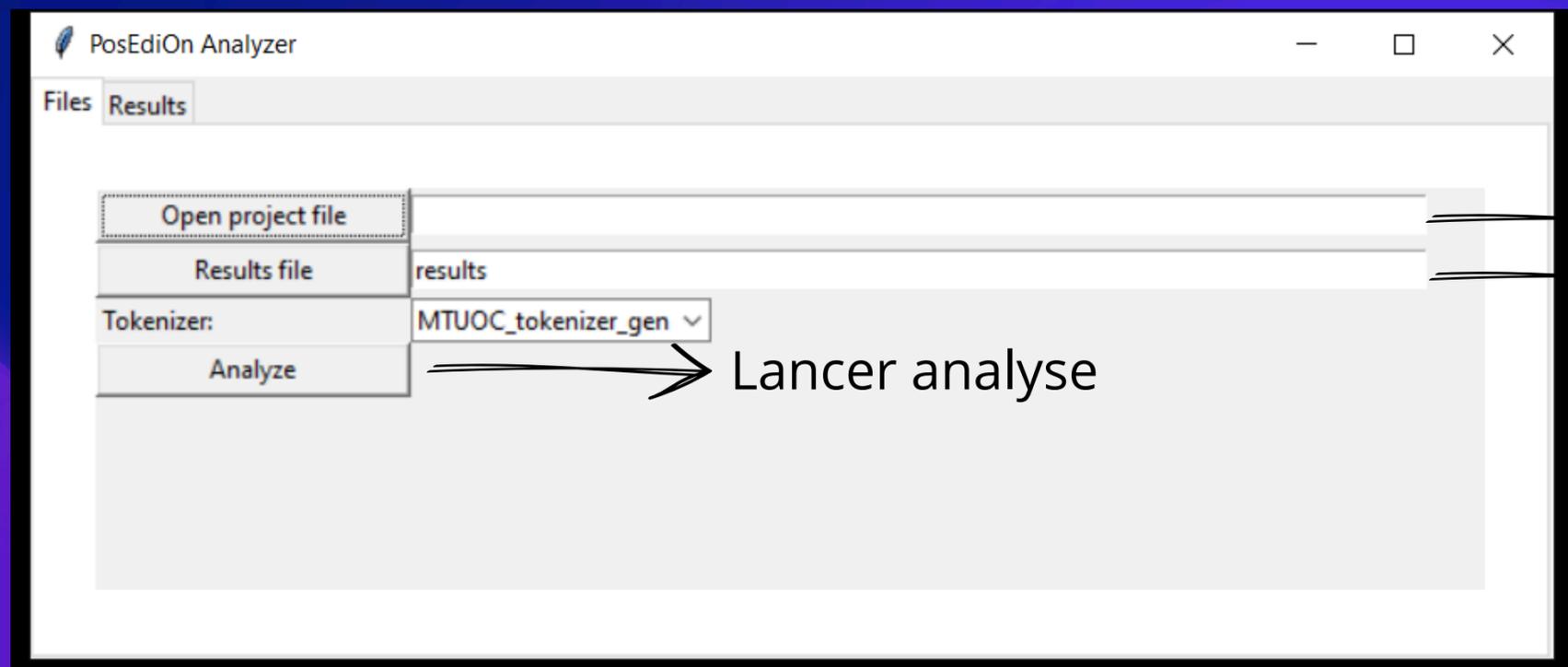
Ils renvoient le fichier .sqlite uniquement



COMPOSANTES : 4-ANALYSE DES DONNÉES

PosEdiOn Analyzer permet d'extraire les données de processus du fichier .sqlite

Format Excel + visualisation dans le programme directement

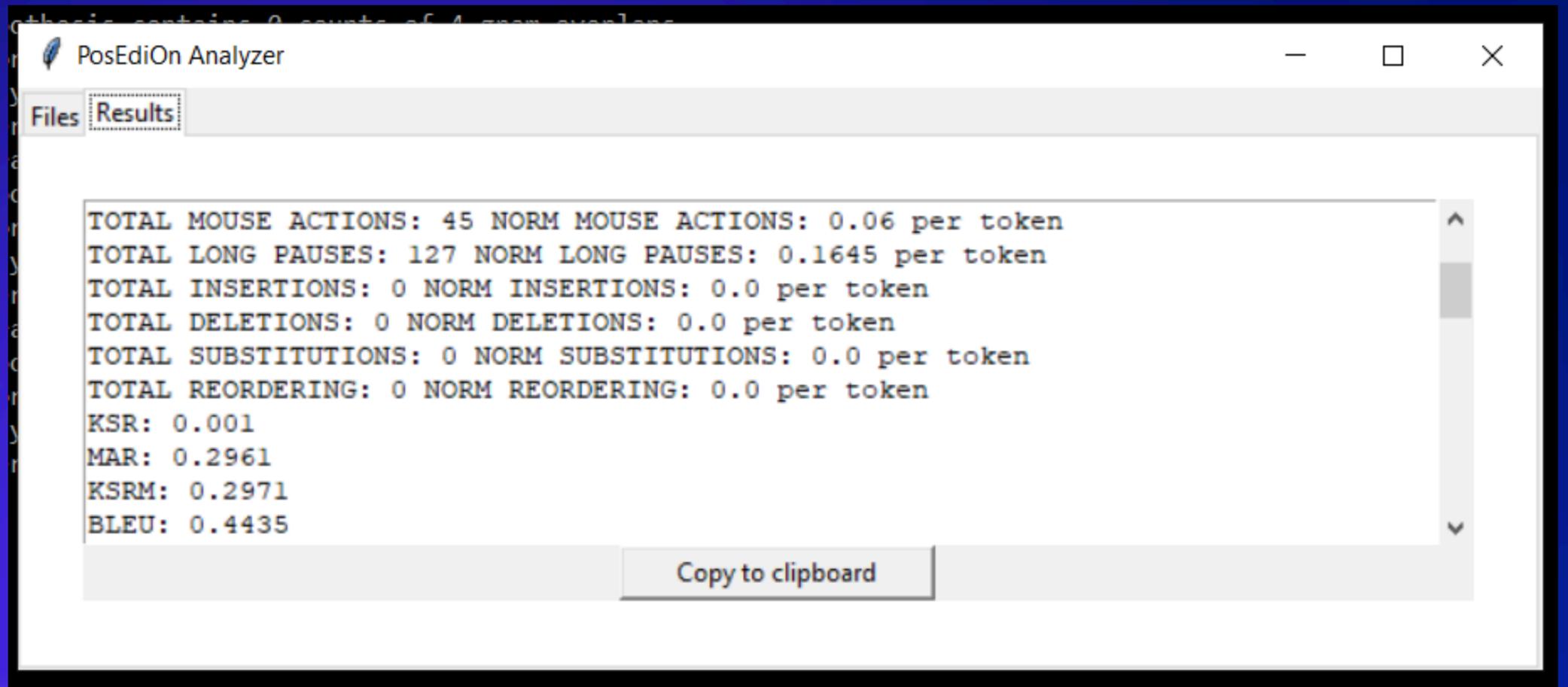


Ouvrir ici le .sqlite renvoyé par les post-éditeurs
Sélectionner empl. fichier de résultats

Lancer analyse

COMPOSANTES : 4-ANALYSE DES DONNÉES

	A	B	C	D
1	EVALUATED SEGMENTS	42		
2	EQUAL SEGMENTS	4		
3	% EQUAL SEGMENTS	9,5238		
4	TOTAL TOKENS	743		
5	EQUAL TOKENS	25		
6	% EQUAL TOKENS	3,3647		
7	TOTAL TIME	0:09:21		
8	NORM. TIME	0,75	per token	
9	TOTAL KEYSTROKES	98		
0	NORM. KEYSTROKES	0,13	per token	
1	TOTAL MOUSE ACTIONS	79		
2	NORM. MOUSE ACTION:	0,11	per token	
3	TOTAL LONG PAUSES	213		
4	NORM. TOTAL LONG PAU	0,2867	per token	
5	INSERTIONS	16		
6	NORM. INSERTIONS	0,0215	per token	
7	DELETIONS	0		
8	NORM. DELETIONS	0	per token	
9	SUBSTITUTIONS	0		
0	NORM. SUBSTITUTIONS	0	per token	
1	REORDERING	0		
2	NORM. REORDERING	0	per token	
3	KSR	0,033		
4	MAR	0,4907		
5	KSRM	0,5237		
6	BLEU	0,4854		
7	NIST	6,0308		
8	TER	0,37		
9	Ins	61		
0	NORM. Ins	0,0821	per token	
1	Del	39		
2	NORM. Del	0,0525	per token	
3	Sub	148		
4	NORM. Sub	0,1992	per token	



The screenshot shows the 'PosEduOn Analyzer' application window. The 'Results' tab is active, displaying a list of metrics in a monospaced font. A 'Copy to clipboard' button is located at the bottom right of the results area.

```
TOTAL MOUSE ACTIONS: 45 NORM MOUSE ACTIONS: 0.06 per token
TOTAL LONG PAUSES: 127 NORM LONG PAUSES: 0.1645 per token
TOTAL INSERTIONS: 0 NORM INSERTIONS: 0.0 per token
TOTAL DELETIONS: 0 NORM DELETIONS: 0.0 per token
TOTAL SUBSTITUTIONS: 0 NORM SUBSTITUTIONS: 0.0 per token
TOTAL REORDERING: 0 NORM REORDERING: 0.0 per token
KSR: 0.001
MAR: 0.2961
KSRM: 0.2971
BLEU: 0.4435
```



BIBLIOGRAPHIE

Oliver, A., Álvarez, S. & Badia, T. (2020). PosEdiOn: Post-editing assessment in PythOn. In: *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, 403-410.

Krings, H. & Koby, G. (2001). *Repairing texts: Empirical investigations of machine translation post-editing processes*. Vol. 5. Kent, Ohio: Kent State University Press.



THANK YOU!

