

How automatic chunking of real-time recorded writing can contribute to segmentation analysis?

Iris Eshkol-Taravella

Georgeta Cislaru

Amandine Jouvenel

Kehina Manseri



Pro-Text

ANR Pro-TEXT – Les processus de textualisation: modélisations linguistiques, psycholinguistiques et d'apprentissage automatique

Processes of Textualization: Linguistic, Psycholinguistic, and Machine Learning Modeling

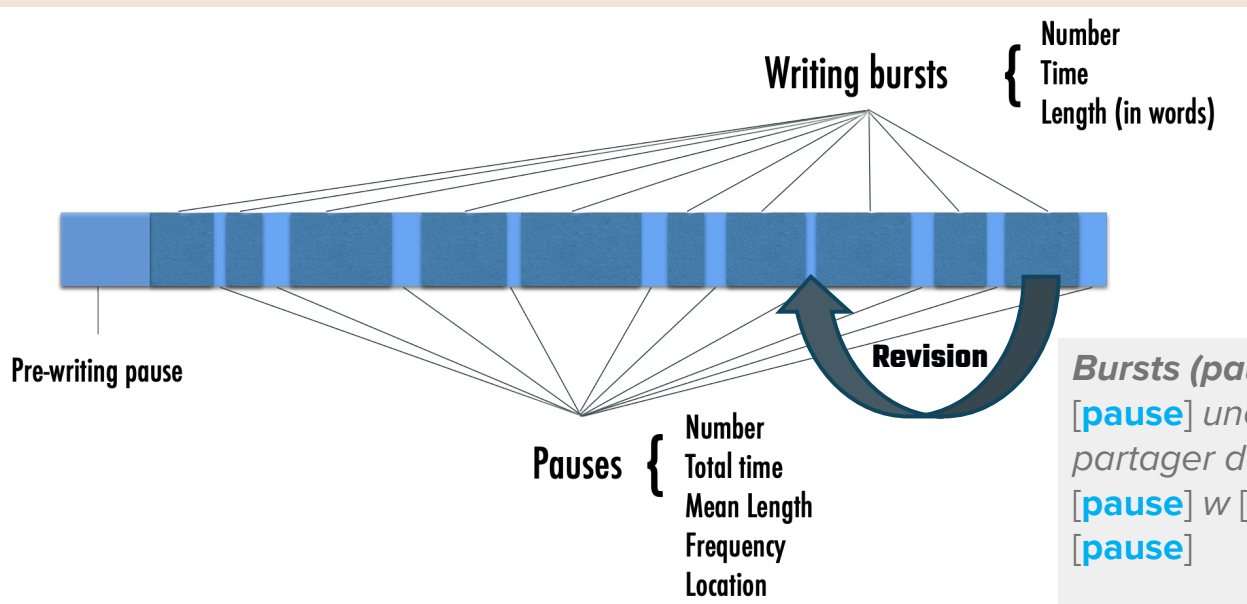
CONTACT

[Georgeta Cislaru](#)

ANR



The writing process is a flow interrupted by pauses & writing events



Pauses are due to:

1. Short pauses
→ Mechanic constraints
2. Longer pauses
→ Cognitive constraints

Bursts (pause threshold: 2 seconds)

[pause] une cousine qui [pause] peut venir partager du temps avec elle pendant [pause] le [pause] w [pause] eek [pause] – [pause] end. [pause]

[pause] a cousin who [pause] can come and spend some time with her during [pause] the [pause] w [pause] eek [pause] – [pause] end. [pause]

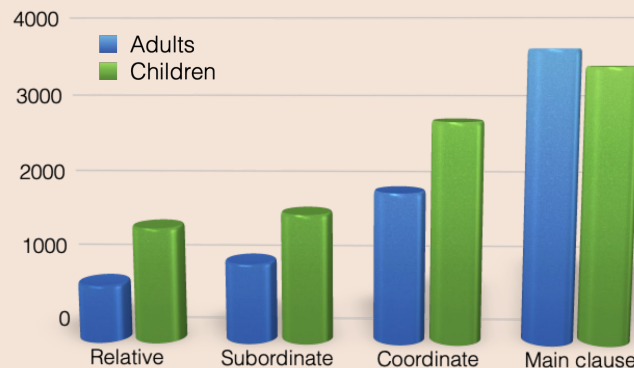
Pauses, a behavioral counterpart of cognitive processes

- **Paragraph** = conceptual planification
- **Sentence or proposition** = syntactical planification
- **Word** = lexical selection, orthograph
- **Within the word** = orthograph, morphological processing, motor skills

Medimorec & Risko (2017)

| Text Boundary | Pause interval (ms) | | |
|---------------|---------------------|-----------|-----------|
| | 300–999 | 1000–1999 | >2000 |
| Word | .44 (.20) | .10 (.06) | .10 (.06) |
| Sentence | .61 (.27) | .15 (.13) | .14 (.16) |
| Paragraph | .89 (.61) | .39 (.41) | .27 (.30) |

van Hell & Verhoeven (2008)



Still searching to understand why writers stop where they stop: from pausology to linguistics

- **Hypothesis 1.**

- Pauses are not random events: they are cognitively but also linguistically meaningful events

- **Hypothesis 2.**

- Language sequences segmented by pauses are relevant linguistic units

- **Analyzing pause location:**

1. Are pauses related to the structural characteristics of texts?
2. Why do pauses appear in certain places in the language flow?
3. Are pauses more frequent at certain linguistic boundaries?
4. To what extent the duration of a pause is related to its location?

A linear model: Towards a sequential approach

- Stage 1 of the “sausage machine”: alignment of about 6 words in the same string (PPP by Frazier & Fodor 1978)
- Grammars of speaking (Blanche-Benveniste 1990, Berrendonner 1994, Brazil 1995, O'Grady 2011)
- Linear Unit Grammar, a unique method of processing speech, writing, reading and listening through small units of meaning called "chunks" (Sinclair and Mauraenen 2006)
- Dynamics and segmentation of information flow in speech and reading (Chafe 1992)
- Preference for a uniform distribution of information (Jaeger 2010)
- Sequential approach: Hierarchical relationships are not necessary for processing (Frank et al. 2012)

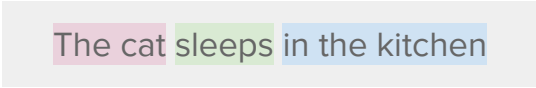
Chunking as a linear model of language processing

- A chunk can be viewed as a computational procedure that *maps, in either direction*, between individual words and sequences of words (Christiansen & Chater, in press). The same viewpoint applies *across linguistic levels*. From this perspective, the chunkatory is not an abstract dictionary of words and phrases with their phonological and semantic properties. Instead, it is a set of bidirectional, incremental procedures, which map between lower level units and chunks composed of sequences of those units. Similarly, the grammar of the language can be viewed as nothing more than a set of incremental procedures mapping, in both directions, between sequences of linguistic units and semantic representation (see, e.g., O'Grady, 2013; Steedman, 2000). A chunk created in comprehension is immediately available for use in production, and vice versa. There is *no need for abstract knowledge of the language, aside from the linguistic structure that is generated through chunking operations*. (Chater et al. 2016: 251)

Chunking

- **Chunk**

- a non-recursive constituent of linguistic units (Abney 1991).



The cat sleeps in the kitchen

- **Chunking (or shallow parsing)**

- identifies *constituents of the sentence* without specifying their internal structure and their syntactic function
- can be done *automatically*
- is based on *previous morphosyntactic labeling*

- **Spontaneous data**

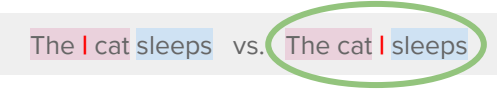
- Chunkers are well adapted for transcribed *oral data* (Tellier et al., 2014; Eshkol-Taravella 2020) : the reparation markers often occur in boundaries of chunks in spoken data (Blanche-Benveniste 1997)
- The chunk is supposed to be a *relevant unit for spontaneous speech* – and probably of language processing more generally.
- Chunking is a good way to perform *syntactic analysis automatically on spontaneous data*, for which full syntactic parsing is often a complex task.

Hypothesis

- General hypothesis:

- *Pausal segmentation* would tend to correspond to *chunk segmentation*

The | cat sleeps vs. The cat | sleeps




Hypothesis

- General hypothesis:

- *Pausal segmentation* would tend to correspond to *chunk segmentation*
- But other phenomena also come into consideration!


The | cat sleeps vs. The cat | sleeps



Hypothesis

- **General hypothesis:**

- *Pausal segmentation* would tend to correspond to *chunk segmentation*
- But other phenomena also come into consideration!




The | cat sleeps vs. The cat | sleeps

- **Chunk segmentation and real-time recorded writing**

Hypothesis

- **General hypothesis:**

- *Pausal segmentation* would tend to correspond to *chunk segmentation*
- But other phenomena also come into consideration!




The | cat sleeps vs. The cat | sleeps

- **Chunk segmentation** and real-time recorded writing

Hypothesis

- **General hypothesis:**

- *Pausal segmentation* would tend to correspond to *chunk segmentation*
- But other phenomena also come into consideration!



The | cat sleeps vs. The cat | sleeps

The diagram shows two versions of the sentence 'The cat sleeps'. In the first version, 'The' is in a pink box and 'cat' is in a blue box, with a vertical red line between them. In the second version, 'The cat' is in a pink box and 'sleeps' is in a blue box, with a vertical red line between them. The second version is circled in green.

- **Chunk segmentation** and real-time recorded writing

→ boundaries

Hypothesis

- **General hypothesis:**

- *Pausal segmentation* would tend to correspond to *chunk segmentation*
- But other phenomena also come into consideration!

The | cat sleeps vs. The cat | sleeps

- **Chunk segmentation** and real-time recorded writing

- boundaries
- category

Hypothesis

- **General hypothesis:**

- *Pausal segmentation* would tend to correspond to *chunk segmentation*
- But other phenomena also come into consideration!

The | cat sleeps vs. The cat | sleeps

- **Chunk segmentation and real-time recorded writing**

- boundaries
- category

Hypothesis

- **General hypothesis:**

- *Pausal segmentation* would tend to correspond to *chunk segmentation*
- But other phenomena also come into consideration!

The | cat sleeps vs. The cat | sleeps

- **Chunk segmentation and real-time recorded writing**

- boundaries
- category

→ pauses

Hypothesis

- General hypothesis:

- *Pausal segmentation* would tend to correspond to *chunk segmentation*
- But other phenomena also come into consideration!

The | cat sleeps vs. The cat | sleeps

- Chunk segmentation and **real-time recorded writing**

→ boundaries
→ category

→ pauses
→ revisions

RB (= Edge revision burst)

In the context of traditional medicine, many people are turning away from old methods. Behind the progress of this new medicine **there are some**

R (= Revision burst)

In the context of traditional medicine **and alternative medicine**, many people are turning away from old methods. Behind the progress of this new medicine there are some

Hypothesis

- Do letters added later in the text appear more often in a certain type of chunk or part of speech?

The cats sleep well vs. The cat sleeps well (Revision burst)

- Do deletions appear more often in a certain type of chunk or part of speech?

The cat | ~~dog~~ sleeps vs. The cat sleeps | ~~eats~~ (Edge revision burst)

RB (= Edge revision burst)

In the context of traditional medicine, many people are turning away from old methods. Behind the progress of this new medicine there are some

R (= Revision burst)

In the context of traditional medicine and alternative medicine, many people are turning away from old methods. Behind the progress of this new medicine there are some

Data (process and product) collected via the keylogger Inputlog (Leijten & Van Waes 2013)

- **Short texts on a given topic with and without production constraints**
 - 165 texts
 - 63 533 words
- **BA psychology students – expert writers**

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_effet_fort_efficace_sur_certaines_points. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchié plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchié_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances au final, les remèdes naturels sont en effet énormément mis en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même et pouvant créer d'autre problèmes médicaux plus ou moins graves chez le plus patient. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances au final, les remèdes naturels sont en effet énormément mis en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même et pouvant créer d'autre problèmes médicaux plus ou moins graves chez le plus patient. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_☒_effet_fort_efficace_sur_certaines_points_☒_☒. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ☒ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<-t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<-t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_effet_fort_efficace_sur_certaines_points. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ecine_alternative | 290 | 306 | 306 |

Preprocessing

csv file

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<-t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<-t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_☒_effet_fort_efficace_sur_certaines_points_☒_☒. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ☒ecine_alternative | 290 | 306 | 306 |

Preprocessing

csv file

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<-t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<-t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_☒_effet_fort_efficace_sur_certaines_points_☒_☒. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ☒ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_effet_fort_efficace_sur_certaines_points. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_effet_fort_efficace_sur_certaines_points. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances (au final), les remèdes naturels sont en effet énormément mis en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même et pouvant créer d'autre problèmes médicaux (plus ou moins grave) chez le plus patient. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances (au final), les remèdes naturels sont en effet énormément mis en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même et pouvant créer d'autre problèmes médicaux (plus ou moins grave) chez le plus patient. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_effet_fort_efficace_sur_certaines_points. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, | les remèdes nature{ }s sont en effet| t énormément mi<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace ut ilise pour la plupart du temps des médicament-s plus ou moins chimi que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer |d'autre problème-s mé-di-caux{ plus ou moins grave} chez le plus patient}. |Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, | les remèdes nature{ }s sont en effet| t énormément mi<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace ut ilise pour la plupart du temps des médicament-s plus ou moins chimi que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer |d'autre problème-s mé-di-caux{ plus ou moins grave} chez le plus patient}. |Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_effet_fort_efficace_sur_certaines_points. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_effet_fort_efficace_sur_certaines_points. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances{ au final}, les remèdes naturels{ } sont en effet énormément mis<t> en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer d'autre problèmes médicaux{ plus ou moins grave} chez le plus patient}. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème *

*Final text **without** annotations of writing phenomena*

** Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted.
However, alternative medicine can pose a lot of problem*

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème *

Final text **without** annotations of writing phenomena

** Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problem*

~~~De plus la médecine alternative |n'utilise~~~ pas que les croyances, |les remèdes naturel{~~}s sont en effe|~~~~~t éno~rmément mis en avant. |Cependant, la médecine alternative{ peut} pos<~é> beaucoup de problème

Final text **with** annotations of writing phenomena

# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème \*

Final text **without** annotations of writing phenomena

*\* Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problem*

~~~De plus la médecine alternative | n'utilise~~~ pas que les croyances | les remèdes naturel{~~}s sont en effet | ~~~~~t éno~rmément mis en avant. | Cependant, la médecine alternative{ peut} pos<~é> beaucoup de problème

Pauses

Final text **with** annotations of writing phenomena

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème *

Final text **without** annotations of writing phenomena

** Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problem*

~~~De plus la médecine alternative | n'utilise~~~ pas que les croyances | les remèdes naturels {~~} sont en effet | ~~~~~t énormément mis en avant. | Cependant, la médecine alternative { peut } pos<~é> beaucoup de problème

**Pauses**      Revision of multiple characters (**deletion**, **insertion**)

Final text **with** annotations of writing phenomena

# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème \*

Final text **without** annotations of writing phenomena

*\* Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problem*

~~~De plus la médecine alternative | n'utilise~~~ pas que les croyances | les remèdes naturels {~~} sont en effet | ~~~~~t énormément mis en avant. | Cependant, la médecine alternative { peut } pos<~é> beaucoup de problème

Pauses

Revision of multiple characters (deletion, insertion)

Revision of a unique character

Final text **with** annotations of writing phenomena

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème *

Final text **without** annotations of writing phenomena

** Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problem*

De plus la médecine alternative n'utilise pas que les croyances les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

Pauses

Revision of multiple characters (deletion, insertion)

Revision of a unique character

Deleted characters

Final text **with** annotations of writing phenomena

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturels sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup ténom même mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |



Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

| n | charBurst | start Pos | doc Length |
|---|---|-----------|------------|
| 1 | L'uⓧⓧⓧDe plus la médecine alternative n'utilise pasⓧⓧⓧ pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturels sont en effet | 65 | 101 |
| 3 | ⓧⓧ | 86 | 99 |
| 4 | beaucoupⓧⓧⓧⓧⓧⓧⓧⓧt énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | ⓧé | 169 | 190 |



- First, the participant typed:

“L'uⓧⓧⓧDe plus la médecine alternative n'utilise pasⓧⓧⓧ
pas que les croyances,”

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'uⓧⓧⓧDe plus la médecine alternative n'utilise pasⓧⓧⓧ pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturels sont en effet | 65 | 101 |
| 3 | ⓧⓧ | 86 | 99 |
| 4 | beaucoupⓧⓧⓧⓧⓧⓧⓧⓧt énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problèmes | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | ⓧé | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances,

- First, the participant typed:

“L'uⓧⓧⓧDe plus la médecine alternative n'utilise pasⓧⓧⓧ pas que les croyances,”

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances,

| n | charBurst | start Pos | doc Length |
|---|---|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturels sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problèmes | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances,

- First, the participant typed:

“L'u De plus la médecine alternative n'utilise pas pas que les croyances,”

- The **deletion characters** indicate that the sequences “L'u” and “pas” were deleted.
- These deleted characters are **replaced with squiggly lines (“~”)**.

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

~~~De plus la médecine alternative n'utilise~~~ pas que les croyances,

| n | charBurst                                                                                                | start Pos | doc Length |
|---|----------------------------------------------------------------------------------------------------------|-----------|------------|
| 1 | L'u⊠⊠⊠De plus la médecine alternative n'utilise pas⊠⊠⊠ pas que les croyances,                            | 0         | 65         |
| 2 | les remèdes naturels sont en effet                                                                       | 65        | 101        |
| 3 | ⊠⊠                                                                                                       | 86        | 99         |
| 4 | beaucoup⊠⊠⊠⊠⊠⊠⊠⊠⊠t énom⊠ même mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99        | 185        |
| 5 | peut                                                                                                     | 159       | 190        |
| 6 | ⊠é                                                                                                       | 169       | 190        |

De plus la médecine alternative n'utilise pas que les croyances,



# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

~~~De plus la médecine alternative n'utilise~~~ pas que les croyances,

| n | charBurst | start Pos | doc Length |
|---|---|-----------|------------|
| 1 | L'u⊠⊠⊠De plus la médecine alternative n'utilise pas⊠⊠⊠ pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturels sont en effet | 65 | 101 |
| 3 | ⊠⊠ | 86 | 99 |
| 4 | beaucoup⊠⊠⊠⊠⊠⊠⊠⊠⊠t énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problèmes | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | ⊠é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances,

- Then, they typed:

“les remèdes naturels sont en effet”

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

~~~De plus la médecine alternative n'utilise~~~ pas que les croyances,

| n | charBurst                                                                                       | start Pos | doc Length |
|---|-------------------------------------------------------------------------------------------------|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances,                        | 0         | 65         |
| 2 | les remèdes naturels sont en effet                                                              | 65        | 101        |
| 3 |                                                                                                 | 86        | 99         |
| 4 | beaucoup énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problèmes | 99        | 185        |
| 5 | peut                                                                                            | 159       | 190        |
| 6 | é                                                                                               | 169       | 190        |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

- Then, they typed:

“les remèdes naturels sont en effet”

# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

| n | charBurst                                                                                         | start Pos | doc Length |
|---|---------------------------------------------------------------------------------------------------|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances,                          | 0         | 65         |
| 2 | les remèdes naturelles sont en effet                                                              | 65        | 101        |
| 3 |                                                                                                   | 86        | 99         |
| 4 | beaucoup t énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problèmes | 99        | 185        |
| 5 | peut                                                                                              | 159       | 190        |
| 6 | é                                                                                                 | 169       | 190        |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

- Then, they typed:

“les remèdes naturelles sont en effet”

- The burst is separated from the existing text by a pipe (“|”) = pause

# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

| n | charBurst                                                                                         | start Pos | doc Length |
|---|---------------------------------------------------------------------------------------------------|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances,                          | 0         | 65         |
| 2 | les remèdes naturels sont en effet                                                                | 65        | 101        |
| 3 |                                                                                                   | 86        | 99         |
| 4 | beaucoup ténom même nt mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99        | 185        |
| 5 | peut                                                                                              | 159       | 190        |
| 6 | é                                                                                                 | 169       | 190        |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

| n | charBurst                                                                                         | start Pos | doc Length |
|---|---------------------------------------------------------------------------------------------------|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances,                          | 0         | 65         |
| 2 | les remèdes naturels sont en effet                                                                | 65        | 101        |
| 3 |                                                                                                   | 86        | 99         |
| 4 | beaucoup ténom même nt mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99        | 185        |
| 5 | peut                                                                                              | 159       | 190        |
| 6 | é                                                                                                 | 169       | 190        |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet



- After a pause, the participant placed their cursor after the **86th** character

# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

~~~De plus la médecine alternative n'utilise~~~ pas que les croyances, |les remèdes naturelles sont en effet

| n | charBurst | start Pos | doc Length |
|---|---|-----------|------------|
| 1 | L'uⓧⓧⓧDe_plus_la_médecine_alternative_n'utilise pasⓧⓧⓧ_pas_que_les_croyances, | 0 | 65 |
| 2 | les_remedes_naturelles_sont_en_effet | 65 | 101 |
| 3 | ⓧⓧ | 86 | 99 |
| 4 | _beaucoupⓧⓧⓧⓧⓧⓧⓧⓧ_t_énomⓧ_même mis_en_avant_Cependant_la_médecine_alternative_pose_beaucoup_de_problème | 99 | 185 |
| 5 | _peut | 159 | 190 |
| 6 | ⓧé | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturelles sont en effet



- After a pause, the participant placed their cursor after the **86th** character, and typed two **deletion characters**: “ⓧⓧ” deleting the sequence “le” from the word “naturelles”

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

- After a pause, the participant placed their cursor after the 86th character, and typed two **deletion characters**: “`<x<x`” deleting the sequence “le” from the word “naturelles”

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

| n | charBurst | start Pos | doc Length |
|---|---|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problèmes | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

- After a pause, the participant placed their cursor after the 86th character, and typed two **deletion characters**: “`<x<x`” deleting the sequence “le” from the word “naturelles”
- These two deleted characters are **replaced with squiggly lines (“~”)**.
- They are also **surrounded by curly brackets (“{ }”)**, indicating that the modification (here, the **deletion**) occurred within the existing text

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t'énom même nt mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t'énom même nt mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

- They then moved their cursor to the end of the text

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t'énom même nt mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

- They then moved their cursor to the **end** of the text and typed:


```

" beaucoup t'énom même nt mis en avant. Cependant,
la médecine alternative pose beaucoup de problème"
            
```

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturels sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t'énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problèmes

- They then moved their cursor to the **end** of the text and typed:


```
" beaucoup t'énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème"
```

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

~~~De plus la médecine alternative n'utilise~~~ pas que les croyances, |les remèdes naturel{~~}s sont en effet |~~énormément~~ énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème

| n | charBurst                                                                                                                 | start Pos | doc Length |
|---|---------------------------------------------------------------------------------------------------------------------------|-----------|------------|
| 1 | L'u⊠⊠⊠De_⊠plus_⊠la_⊠médecine_⊠alternative_⊠n'utilise pas_⊠⊠⊠_⊠pas_⊠que_⊠les_⊠croyances,_⊠                                 | 0         | 65         |
| 2 | les_⊠remèdes_⊠naturelles_⊠s ont_⊠en_⊠effet                                                                                | 65        | 101        |
| 3 | ⊠⊠                                                                                                                        | 86        | 99         |
| 4 | ⊠_⊠beaucoup_⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠t_⊠énom⊠rmément_⊠mis_⊠en_⊠avant_⊠Cependant_⊠la_⊠médecine_⊠alternative_⊠pose_⊠beaucoup_⊠de_⊠problème | 99        | 185        |
| 5 | ⊠_⊠peut                                                                                                                   | 159       | 190        |
| 6 | ⊠é                                                                                                                        | 169       | 190        |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème

- They then moved their cursor to the end of the text and typed:
  - “⊠\_⊠beaucoup\_⊠⊠⊠⊠⊠⊠⊠⊠⊠⊠t\_⊠énom⊠rmément\_⊠mis\_⊠en\_⊠avant\_⊠Cependant,
  - ⊠\_⊠la\_⊠médecine\_⊠alternative\_⊠pose\_⊠beaucoup\_⊠de\_⊠problème”
- The deleted characters are **replaced with squiggly lines (“~”)**.
- The burst is **separated from the existing text by a pipe (“|”) = pause**

# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

*Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems*

~~~De plus la médecine alternative n'utilise~~~ pas que les croyances, |les remèdes naturel{~~}s sont en effe|~~~~~t éno~rmément mis en avant. Cependant, la médecine alternative pose beaucoup de problème

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u⊠⊠⊠De plus la médecine alternative n'utilise pas⊠⊠⊠ pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | ⊠⊠ | 86 | 99 |
| 4 | beaucoup⊠⊠⊠⊠⊠⊠⊠⊠⊠t énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | ⊠é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème

Reconstructed texts with conservation of writing phenomena


De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problèmes

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problèmes

| n | charBurst | start Pos | doc Length |
|---|---|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturels sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup ténom même nt mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problèmes



- The participant placed their cursor after the 159th character



Reconstructed texts with conservation of writing phenomena


De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème


Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème

| n | charBurst | start Pos | doc Length |
|---|---|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup ténom même nt mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème



- The participant placed their cursor after the 159th character and typed: “ `peut`”
- 

Reconstructed texts with conservation of writing phenomena


De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème


Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

~~~De plus la médecine alternative n'utilise~~~ pas que les croyances, |les remèdes naturel{~~}s sont en effe|~~~~~t éno~rmément mis en avant. Cependant, la médecine alternative pose beaucoup de problème

| n | charBurst                                                                                                 | start Pos | doc Length |
|---|-----------------------------------------------------------------------------------------------------------|-----------|------------|
| 1 | L'u⊠⊠⊠De plus la médecine alternative n'utilise pas⊠⊠⊠ pas que les croyances,                             | 0         | 65         |
| 2 | les remèdes naturels sont en effet                                                                        | 65        | 101        |
| 3 | ⊠⊠                                                                                                        | 86        | 99         |
| 4 | beaucoup⊠⊠⊠⊠⊠⊠⊠⊠⊠t énorm⊠ même mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99        | 185        |
| 5 | peut                                                                                                      | 159       | 190        |
| 6 | ⊠é                                                                                                        | 169       | 190        |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut pose beaucoup de problème



- The participant placed their cursor after the 159th character and typed: “ peut”
- 

# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative **{ peut }** pose beaucoup de problème

| n | charBurst                                                                                        | start Pos | doc Length |
|---|--------------------------------------------------------------------------------------------------|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances,                         | 0         | 65         |
| 2 | les remèdes naturels sont en effet                                                               | 65        | 101        |
| 3 |                                                                                                  | 86        | 99         |
| 4 | beaucoup t énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99        | 185        |
| 5 | peut                                                                                             | 159       | 190        |
| 6 | é                                                                                                | 169       | 190        |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative **peut** pose beaucoup de problème

- The participant placed their cursor after the 159th character and typed: “ **peut**”
- In the reconstructed text, this sequence of characters is **surrounded by curly brackets (“{ }”)**, indicating that the modification (in this case, the **insertion**) occurred within the existing text.

# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

~~~De plus la médecine alternative n'utilise~~~ pas que les croyances, |les remèdes naturel{~~}s sont en effe|~~~~~t éno~rmément mis en avant. Cependant, la médecine alternative{ peut} pose beaucoup de problème

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u⊠⊠⊠De plus la médecine alternative n'utilise pas⊠⊠⊠ pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | ⊠⊠ | 86 | 99 |
| 4 | beaucoup⊠⊠⊠⊠⊠⊠⊠⊠⊠t énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | ⊠é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut pose beaucoup de problème

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problème

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturels sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t'énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut pose beaucoup de problème



- Finally, the participant placed their cursor after the 169th character

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut poser beaucoup de problème

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturels sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t'énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème



- Finally, the participant placed their cursor after the 169th character and typed: “`<x>é`”
- → deleting the “e” from “pose” and replacing it with “é”.

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut pose beaucoup de problème

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t'énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

- Finally, the participant placed their cursor after the 169th character and typed: “`<x>é`”
- → deleting the “e” from “pose” and replacing it with “é”.

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

| n | charBurst | start Pos | doc Length |
|---|--|-----------|------------|
| 1 | L'u De plus la médecine alternative n'utilise pas pas que les croyances, | 0 | 65 |
| 2 | les remèdes naturelles sont en effet | 65 | 101 |
| 3 | | 86 | 99 |
| 4 | beaucoup t'énormément mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99 | 185 |
| 5 | peut | 159 | 190 |
| 6 | é | 169 | 190 |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

- Finally, the participant placed their cursor after the 169th character and typed: “<é”
- → deleting the “e” from “posé” and replacing it with “é”.
- The deleted “e” is replaced with a squiggly line (“~”).
- The burst is surrounded by chevrons (“<>”), indicating it is a single-character insertion within the existing text.

Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problems

~~~De plus la médecine alternative n'utilise~~~ pas que les croyances, |les remèdes naturel{~~}s sont en effe|~~~~~t éno~rmément mis en avant. Cependant, la médecine alternative{ peut} pos<~é> beaucoup de problème

| n | charBurst                                                                                               | start Pos | doc Length |
|---|---------------------------------------------------------------------------------------------------------|-----------|------------|
| 1 | L'u⊠⊠⊠De plus la médecine alternative n'utilise pas⊠⊠⊠ pas que les croyances,                           | 0         | 65         |
| 2 | les remèdes naturelles sont en effet                                                                    | 65        | 101        |
| 3 | ⊠⊠                                                                                                      | 86        | 99         |
| 4 | beaucoup⊠⊠⊠⊠⊠⊠⊠⊠t énom⊠ même mis en avant. Cependant, la médecine alternative pose beaucoup de problème | 99        | 185        |
| 5 | peut                                                                                                    | 159       | 190        |
| 6 | ⊠é                                                                                                      | 169       | 190        |

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème



# Reconstructed texts with conservation of writing phenomena

De plus la médecine alternative n'utilise pas que les croyances, les remèdes naturels sont en effet énormément mis en avant. Cependant, la médecine alternative peut posé beaucoup de problème

*Final text **without** annotations of writing phenomena*

*\* Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problem*

# Reconstructed texts with conservation of writing phenomena

~~~De plus la médecine alternative | n'utilise~~~ pas que les croyances, | les remèdes naturel{~~~}s sont en effet|~~~~~t énormément mis en avant. | Cependant, la médecine alternative{ peut} pos<~é> beaucoup de problème

Final text **with** annotations of writing phenomena

* *Moreover, alternative medicine does not only use beliefs, natural remedies are in fact greatly promoted. However, alternative medicine can pose a lot of problem*

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_effet_fort_efficace_sur_certaines_points. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances au final, les remèdes naturels sont en effet énormément mis en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même et pouvant créer d'autre problèmes médicaux plus ou moins graves chez le plus patient. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena

Chunking

De plus nous pouvons souligner le fait que la médecine alternative n'utilise pas que les croyances au final, les remèdes naturels sont en effet énormément mis en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimiques que pouvant parfois être plus nocif que la maladie en elle-même et pouvant créer d'autre problèmes médicaux plus ou moins graves chez le plus patient. Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Data processing

| ID | n_burst | categ | burst | charBurst | startPos | endPos | docLength |
|------|---------|-------|--|--|----------|--------|-----------|
| P+S1 | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 |
| P+S1 | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme. | 45 | 76 | 76 |
| P+S1 | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_☒_effet_fort_efficace_sur_certaines_points_☒_☒. | 76 | 170 | 170 |
| P+S1 | 4 | P | En effet si l'ont y réfléchie plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchie_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 |
| P+S1 | 5 | RB | ecine alternative | ☒ecine_alternative | 290 | 306 | 306 |

csv file

Preprocessing

~~~De plus nous pouvons ~~~souligner le fait que la médecine alternative |n'utilise~~~ pas que les croyances{ au final}|, |les remèdes naturel{~~~}s sont en effet|~~~~~t éno~rmément mi<-t> en avant contrairement à la médecine actuelle qui mêm~e si elle s'avère efficace ut~~~~~ilise pour la plupart du~ temps des médicament~s plus ou moins chimi~~~~~que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer |d'autre problème~s mé~di~~~~caux{ plus ou moins grave} chez le plus patient}. |Cependant, si nous devons faire une comparaison entre ces

Chunking

texts with conservation of writing phenomena

~~~De plus nous pouvons ~~~souligner le fait que la médecine alternative |n'utilise~~~ pas que les croyances{ au final}|, |les remèdes naturel{~~~}s sont en effet|~~~~~t éno~rmément mi<-t> en avant contrairement à la médecine actuelle qui mêm~e si elle s'avère efficace ut~~~~~ilise pour la plupart du~ temps des médicament~s plus ou moins chimi~~~~~que pouvant parfois être plus nocif que la maladie en elle-même{ et pouvant créer |d'autre problème~s mé~di~~~~caux{ plus ou moins grave} chez le plus patient}. |Cependant, si nous devons faire une comparaison entre ces

texts with conservation of writing phenomena and chunk segmentation

Chunking of reconstructed texts

Adding another segmentation layer to reconstructed texts

Chunking of reconstructed texts

Adding another segmentation layer to reconstructed texts

Level 1: Pauses (writing phenomena)

Depuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes passent beaucoup de temps | sur le téléphone portable ainsi que sur | leur écran

Chunking of reconstructed texts

Adding another segmentation layer to reconstructed texts

Level 1: Pauses (writing phenomena)

Depuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes passent beaucoup de temps | sur le téléphone portable ainsi que sur | leur écran

Level 2: Errors (writing phenomena)

D[~]epuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes {passent} beaucoup de temps | sur le téléphone portable ainsi q<u>e sur | leur écran

[~] character deletion

{ } characters sequence insertion

<> single character insertion

Chunking of reconstructed texts

Adding another segmentation layer to reconstructed texts

Level 1: Pauses (writing phenomena)

Depuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes passent beaucoup de temps | sur le téléphone portable ainsi que sur | leur écran

Level 2: Errors (writing phenomena)

D[~]epuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes {passent} beaucoup de temps | sur le téléphone portable ainsi q<u>e sur | leur écran

[~] character deletion

{ } characters sequence insertion

< > single character insertion

Level 3: Chunks

[D[~]epuis l'avancée] [des nouvelles technologies] [et] | [l'hyperconnectivité] [qui] [en] [est] | [le résultat] [les jeunes] [{passent}] [beaucoup] [de temps] | [sur le téléphone portable] [ainsi q<u>e] [sur | leur écran]

PP NP CONJ AdP VN

Chunking of reconstructed texts

Adding another segmentation layer to reconstructed texts

Level 1: Pauses (writing phenomena)

Depuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes passent beaucoup de temps | sur le téléphone portable ainsi que sur | leur écran

Level 2: Errors (writing phenomena)

D[~]epuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes {passent} beaucoup de temps | sur le téléphone portable ainsi q<u>e sur | leur écran

[~] character deletion

{ } characters sequence insertion

< > single character insertion

Level 3: Chunks

[D[~]epuis l'avancée] [des nouvelles technologies] [et] | [l'hyperconnectivité] [qui] [en] [est] | [le résultat] [les jeunes] [{passent}] [beaucoup] [de temps] | [sur le téléphone portable] [ainsi q<u>e] [sur | leur écran]

PP NP CONJ AdP VN

Mandatory steps

Chunking of reconstructed texts

Adding another segmentation layer to reconstructed texts

Level 1: Pauses (writing phenomena)

Depuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes passent beaucoup de temps | sur le téléphone portable ainsi que sur | leur écran

Level 2: Errors (writing phenomena)

D[~]epuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes {passent} beaucoup de temps | sur le téléphone portable ainsi q<u>e sur | leur écran

[~] character deletion

{ } characters sequence insertion

<u> single character insertion

Level 3: Chunks

[D[~]epuis l'avancée] [des nouvelles technologies] [et] | [l'hyperconnectivité] [qui] [en] [est] | [le résultat] [les jeunes] [{passent}] [beaucoup] [de temps] | [sur le téléphone portable] [ainsi q<u>e] [sur | leur écran]

PP NP CONJ AdP VN

Mandatory steps

- Use a segmentation tool (chunking software)

Chunking of reconstructed texts

Adding another segmentation layer to reconstructed texts

Level 1: Pauses (writing phenomena)

Depuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes passent beaucoup de temps | sur le téléphone portable ainsi que sur | leur écran

Level 2: Errors (writing phenomena)

D[~]epuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes {passent} beaucoup de temps | sur le téléphone portable ainsi q<u>e sur | leur écran

[~] character deletion

{ } characters sequence insertion

<u> single character insertion

Level 3: Chunks

[D[~]epuis l'avancée] [des nouvelles technologies] [et] | [l'hyperconnectivité] [qui] [en] [est] | [le résultat] [les jeunes] [{passent}] [beaucoup] [de temps] | [sur le téléphone portable] [ainsi q<u>e] [sur | leur écran]

PP NP CONJ AdP VN

Mandatory steps

- Use a segmentation tool (chunking software)
- Process the corpus to allow relevant counts

Chunking of reconstructed texts

Adding another segmentation layer to reconstructed texts

Level 1: Pauses (writing phenomena)

Depuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes passent beaucoup de temps | sur le téléphone portable ainsi que sur | leur écran

Level 2: Errors (writing phenomena)

D[~]epuis l'avancée des nouvelles technologies et | l'hyperconnectivité qui en est | le résultat les jeunes {passent} beaucoup de temps | sur le téléphone portable ainsi q<u>e sur | leur écran

[~] character deletion

{ } characters sequence insertion

<u> single character insertion

Level 3: Chunks

[D[~]epuis l'avancée] [des nouvelles technologies] [et] | [l'hyperconnectivité] [qui] [en] [est] | [le résultat] [les jeunes] [{passent}] [beaucoup] [de temps] | [sur le téléphone portable] [ainsi q<u>e] [sur | leur écran]

PP NP CONJ AdP VN

Mandatory steps

- Use a segmentation tool (chunking software)
- Process the corpus to allow relevant counts
- Use those counts to corroborate hypotheses

Chunking of reconstructed texts

Some specificities :

- Contains punctuation marks (errors and pauses)
- Contains typographical errors

Chunking of reconstructed texts

SEM (Dupont & Tellier, 2017)

Syntactic annotation software for french

3 levels of annotations (Part-Of-Speech, chunking and entity recognition)

Chunking of reconstructed texts

SEM (Dupont & Tellier, 2017)

Syntactic annotation software for french

3 levels of annotations (Part-Of-Speech, chunking and entity recognition)

| Types | Details | POS |
|-------------|---------------------|---|
| __UNKNOWN__ | Unidentified chunks | ET |
| VN | Verbal chunks | V, VIMP, VINF, VPP, VPR, VS |
| AP | Adjectival chunks | ADJ, ADJWH |
| AdP | Adverbial chunks | ADV, ADVWH, I |
| CONJ | Conjunction chunks | CC, CS |
| NP | Noun chunks | CLO, CLR, CLS, NC, NPP, PRO, PROREL, PROWH |
| PP | Preposition chunks | P, P+D, P+PRO |

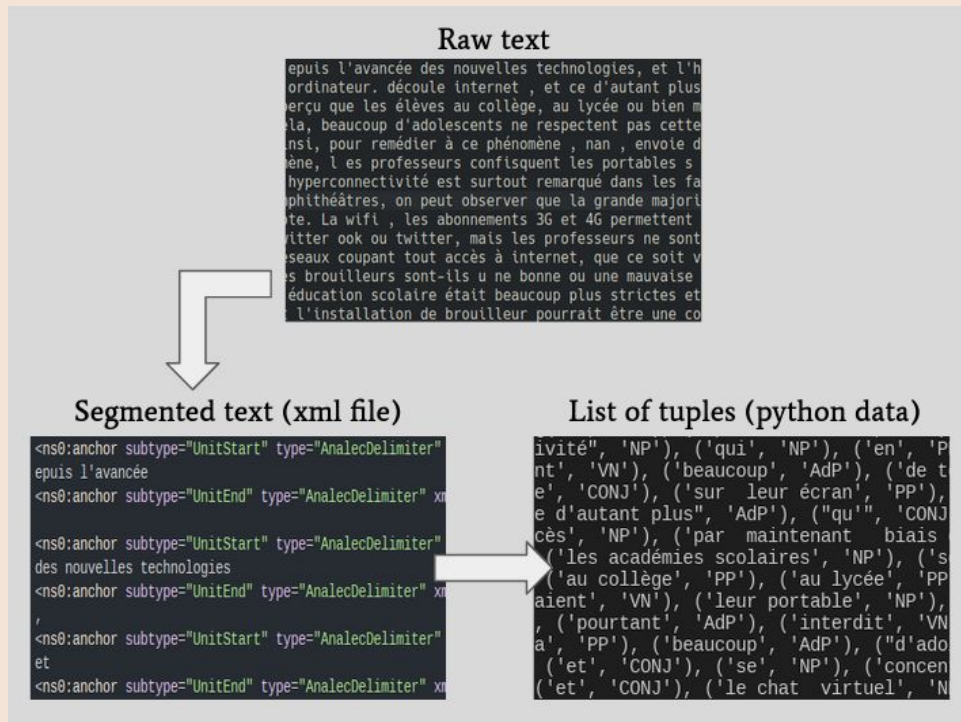
Chunking of reconstructed texts

SEM (Dupont & Tellier, 2017)

Syntactic annotation software for french

3 levels of annotations (Part-Of-Speech, chunking and entity recognition)


| Types | Details | POS |
|-------------|---------------------|--|
| __UNKNOWN__ | Unidentified chunks | ET |
| VN | Verbal chunks | V, VIMP, VINF, VPP, VPR, VS |
| AP | Adjectival chunks | ADJ, ADJWH |
| AdP | Adverbial chunks | ADV, ADVWH, I |
| CONJ | Conjunction chunks | CC, CS |
| NP | Noun chunks | CLO, CLR, CLS, NC, NPP, PRO, PROREL, PROWH |
| PP | Preposition chunks | P, P+D, P+PRO |




Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou



[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
'], 'AdP'), ('[des brouilleurs]',
'PP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
'], 'AdP'), ('[attentifs]', 'AP'),



[('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP'), ('[écouter]',
''), ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
'], 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
'], 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[| tte | décision]', 'NP'), ('[peut]', 'VN'), ('

Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter e
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou

[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
]', 'AdP'), ('[des brouilleurs]',
]', 'AdP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
, 'AdP'), ('[attentifs]', 'AP'),


[('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP'), ('[écouter]',
, ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
, ('[Il]', 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
, ('[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[| tte | décision]', 'NP'), ('[peut]', 'VN'), (

2 types of pauses:


Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter e
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou



[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
'], 'AdP'), ('[des brouilleurs]',
'PP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
'], 'AdP'), ('[attentifs]', 'AP'),



('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP') ('[écouter]',
'[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
'[Il]', 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
'[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[| tte | décision]', 'NP'), ('[peut]', 'VN'), ('

2 types of pauses:

Pauses in-between chunks

Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter e
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou

[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
'], 'AdP'), ('[des brouilleurs]',
'PP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
'], 'AdP'), ('[attentifs]', 'AP'),

(('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP') ('[écouter]',
'V'), ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
'], 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
'], 'CONJ'), ('[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[| tte | décision]', 'NP'), ('[peut]', 'VN'), ('

2 types of pauses:

Pauses in-between chunks

Pauses within chunks

Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter e
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou

[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
]', 'AdP'), ('[des brouilleurs]',
'PP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
, 'AdP'), ('[attentifs]', 'AP'),

(('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP'), ('[écouter]',
'V'), ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
, ('[Il]', 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
, ('[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[tte | décision]', 'NP'), ('[peut]', 'VN'), ('

2 types of pauses:

Pauses in-between chunks


Pauses within chunks

To simplify counts:

Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou



[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
'], 'AdP'), ('[des brouilleurs]',
'PP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
'], 'AdP'), ('[attentifs]', 'AP'),

(('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP'), ('[écouter]',
'V'), ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
'], 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
'], 'CONJ'), ('[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[| tte | décision]', 'NP'), ('[peut]', 'VN'), ('

2 types of pauses:

Pauses in-between chunks

Pauses within chunks

To simplify counts:

- Transform “fake” chunks: ('[|]', 'NP') into ('|', 'PAUSE')

Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter e
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou

[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
]', 'AdP'), ('[des brouilleurs]',
'PP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
, 'AdP'), ('[attentifs]', 'AP'),

(('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP') ('[écouter]',
, ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
, ('[Il]', 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
, ('[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[| tte | décision]', 'NP'), ('[peut]', 'VN'), (

2 types of pauses:

Pauses in-between chunks

Pauses within chunks

To simplify counts:

- Transform “fake” chunks: ('[|]', 'NP') into ('|', 'PAUSE')
- Retrieve frontier pipes to place them into their own elements: ('[| moins]', 'AdP')

Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou

[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
]', 'AdP'), ('[des brouilleurs]',
'PP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
' , 'AdP'), ('[attentifs]', 'AP'),

(('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP'), ('[écouter]',
'V'), ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
' , ('[Il]', 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
' , ('[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[tte | décision]', 'NP'), ('[peut]', 'VN'), ('

2 types of pauses:

Pauses in-between chunks

Pauses within chunks

To simplify counts:

- Transform “fake” chunks: ('[|]', 'NP') into ('|', 'PAUSE')
- Retrieve frontier pipes to place them into their own elements: ('[| moins]', 'AdP')

Resulting structure

[mais]', 'CONJ'), ('|', 'PAUSE'),
'E'), ('[Les élèves]', 'NP'), ('[ont
N'), ('[en place]', 'PP'), ('[de ce
élèves]', 'PP'), ('[plus]', 'AdP')

Pauses in-between
chunks
Pauses within
chunks

Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou

[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
]', 'AdP'), ('[des brouilleurs]',
]', 'PP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
, 'AdP'), ('[attentifs]', 'AP'),

(('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP'), ('[écouter]',
, ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
, ('[Il]', 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
, ('[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[| tte | décision]', 'NP'), ('[peut]', 'VN'), (

2 types of pauses:

Pauses in-between chunks

Pauses within chunks

To simplify counts:

- Transform “fake” chunks: ('[|]', 'NP') into ('|', 'PAUSE')
- Retrieve frontier pipes to place them into their own elements: ('[| moins]', 'AdP')

Resulting structure

[mais]', 'CONJ'), ('|', 'PAUSE'),
(E)', ('[Les élèves]', 'NP'), ('[ont
N)', ('[en place]', 'PP'), ('[de ce
élèves]', 'PP'), ('[plus]', 'AdP')

Pauses in-between
chunks
Pauses within
chunks

- Simplify counting of both types
- Retrieve context of both types

Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter
plu de recherche | | - | mo
DANS LES SALLES DE CLASSES
possible que prochainement
installés dans les salles de
notre faculté | . | Ceci pou

[('[avantage : -]', 'NP'), ('[mieu
DES BROUILLEURS]', 'NP'), ('[DANS
]', 'AdP'), ('[des brouilleurs]',
]', 'PP'), ('[les amphithéâtre]', 'NP
antages]', 'NP'), ('[comme des inc
, 'AdP'), ('[attentifs]', 'AP'),

(('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP'), ('[écouter]',
, ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
, ('[Il]', 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
, ('[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
VN'), ('[| avoir]', 'PP'), ('[tte | décision]', 'NP'), ('[peut]', 'VN'), ('

2 types of pauses:

Pauses in-between chunks

Pauses within chunks

To simplify counts:

- Transform “fake” chunks: ('[|]', 'NP') into ('|', 'PAUSE')
- Retrieve frontier pipes to place them into their own elements: ('[| moins]', 'AdP')

Resulting structure

[mais]', 'CONJ'), ('|', 'PAUSE'),
, ('[Les élèves]', 'NP'), ('[ont
N)', ('[en place]', 'PP'), ('[de ce
élèves]', 'PP'), ('[plus]', 'AdP')

Pauses in-between
chunks
Pauses within
chunks

- Simplify counting of both types
- Retrieve context of both types

est pas fait VN
Left context
chunk + type

PAUSE
Pause (in-between chunks)

avec leur accord PP
Right context
chunk + type

Chunking of reconstructed texts

Combining layers: pauses and chunks

avantage: - | mieux écouter
 plsu de recherche | | - | mo
 DANS LES SALLES DE CLASSES
 possible que prochainement
 installés dans les salles de
 notre faculté | . | Ceci pou

[('[avantage : -]', 'NP'), ('[mieu
 DES BROUILLEURS]', 'NP'), ('[DANS
]', 'AdP'), ('[des brouilleurs]',
 'PP'), ('[les amphithéâtre]', 'NP
 antages]', 'NP'), ('[comme des inc
 ', 'AdP'), ('[attentifs]', 'AP'),



(('[avantage : -]', 'NP'), ('[|]', 'NP'), ('[mieux]', 'AdP'), ('[écouter]',
), ('[| moins]', 'AdP'), ('[de triche |]', 'PP'), ('[|]', 'NP'), ('[DES BRO
 ('[Il]', 'NP'), ('[est]', 'VN'), ('[possible]', 'AP'), ('[que]', 'CONJ'),
 , ('[soit]', 'CONJ'), ('[installés]', 'VN'), ('[dans les salles]', 'PP'), ('
 VN'), ('[| avoir]', 'PP'), ('[tte | décision]', 'NP'), ('[peut]', 'VN'), ('

2 types of pauses:

Pauses in-between chunks

Pauses within chunks

To simplify counts:

- Transform “fake” chunks: ('[|]', 'NP') into ('|', 'PAUSE')
- Retrieve frontier pipes to place them into their own elements: ('[| moins]', 'AdP')

Resulting structure

[mais]', 'CONJ'), ('|', 'PAUSE'),
 E'), ('[Les|élèves]', 'NP'), ('[ont
 N'), ('[en place]', 'PP'), ('[de ce
 élèves]', 'PP'), ('[plus]', 'AdP')

Pauses in-between
 chunks
 Pauses within
 chunks

- Simplify counting of both types
- Retrieve context of both types

est pas fait VN
 Left context
 chunk + type

| PAUSE
 Pause (in-between chunks)

avec leur accord PP
 Right context
 chunk + type

témoigne VN
 Left context
 chunk + type

une étudiante | fumeuse NP
 Pause (within chunks)

leur mégo NP
 Right context
 chunk + type

Chunking of reconstructed texts

Combining layers: errors and chunks

[[('Mise', 'NP'), ('en place', 'NP'), ('nous', 'NP'), ('allons', 'VN'), ('de', 'PP'), ("de l'éducation.", 'PP'), ('moir', 'NP'), ('dans chaque f', 'PP'), ('nous', 'NP'), ('demander', 'PP'), ('', 'AP'), ('|', 'PAUSE'), ('nan', 'PP'), ('Mise en pl~ace| au sein de~s différents |', 'PP'), ('l'installation de fumoirs|. En effet, |', 'PP'), ('esure a été décidé<e> suite au|~x ~~~~~', 'PP'), ('tudiants{ fumeurs} qui espéraient pouvo', 'PP'), ('durant les périodes |plus froides et "au', 'PP'), ('d'int~empéries. . | C|ette initiative n'

('pour mieux', 'PP'), ('réussir', 'VN'), ('par la suite .|', 'PP'), ('[Ces appa-re-ils]', 'PP'), ('s|~', 'PP'), ('[|]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('plus', 'AdP'), ('pronon', 'PP'), ('[~fixé]', 'VN'), ('sur leu{ ~-]', 'PP'), ('[] téléphones', 'NP'), ('et]', 'CONJ'), ('nt]', 'VN'), ('[se sent~ir |surveillés]', 'NP'), ('ou]', 'CONJ'), ('[~empêchés]', 'VN'), ('P)', ('perdre', 'VN'), ('leur autonomie|]', 'NP'), ('et]', 'CONJ'), ('leur plaisir', 'NP')

Chunking of reconstructed texts


Combining layers: errors and chunks

[[('Mise', 'NP'), ('en place',
ous]', 'NP'), ('allons', 'VN')
e', 'PP'), ("de l'éducation.
moir]', 'NP'), ('dans chaque f
, ('nous', 'NP'), ('demander'
, 'AP'), ('|', 'PAUSE'), ('nan

Mise en pl^ace| au sein de^s différents
l'installation de fumoirs|. En effet, |
esure a été décidé<e> suite au|^x ~~~~~
tudiants{ fumeurs} qui espéraient pouvo
durant les périodes |plus froides et "au
d'int^empéries. . | C|ette initiative n'



('pour mieux', 'PP'), ('réussir', 'VN'), ('par la suite .]', 'PP'), ('[Ces appa-re-ils',
s|~ ,]', 'PP'), ('[|]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('[plus]', 'AdP'), ('[pronon
, 'PP'), ('[~fixé]', 'VN'), ('[sur leu{ ~~]', 'PP'), ('[] téléphones]', 'NP'), ('[et]', 'CONJ'),
nt]', 'VN'), ('[se sent~ir |surveillé-s]', 'NP'), ('[ou]', 'CONJ'), ('[|~empêchés]', 'VN'), ('
P'), ('[perdre]', 'VN'), ('[leur autonomie]', 'NP'), ('[et]', 'CONJ'), ('[leur plaisir]', 'NP')



- How can we find errors automatically ?

Chunking of reconstructed texts

Combining layers: errors and chunks

[('Mise', 'NP'), ('en place', 'NP'), ('nous', 'NP'), ('allons', 'VN'), ('de', 'PP'), ('de l'éducation.', 'PP'), ('moir', 'NP'), ('dans chaque f', 'PP'), ('nous', 'NP'), ('demander', 'PP'), ('', 'AP'), ('|', 'PAUSE'), ('nan', 'NP')]

Mise en place| au sein de~s différents
l'installation de fumoirs|. En effet, |~
mesure a été décidé<e> suite au|~x ~~~~~
tudiants{ fumeurs} qui espéraient pouvo
durant les périodes |plus froides et "au
d'int~empéries. . | C|ette initiative n'

('pour mieux', 'PP'), ('réussir', 'VN'), ('par la suite .', 'PP'), ('[Ces appa-re-ils]', 'NP'), ('s|~', 'PP'), ('[|]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('[plus]', 'AdP'), ('[pronon', 'PP'), ('[~fixé]', 'VN'), ('[sur leu{ ~~]', 'PP'), ('[|] téléphones]', 'NP'), ('[et]', 'CONJ'), ('nt]', 'VN'), ('[se sent~ir |surveillé-s]', 'NP'), ('[ou]', 'CONJ'), ('[|~empêchés]', 'VN'), ('P)', ('[perdre]', 'VN'), ('[leur autonomie]', 'NP'), ('[et]', 'CONJ'), ('[leur plaisir]', 'NP')

- How can we find errors automatically ?
- To simplify counts:

Chunking of reconstructed texts

Combining layers: errors and chunks

[('Mise', 'NP'), ('[en place]', 'NP'), ('[ous]', 'NP'), ('[allons]', 'VN'), ('[e]', 'PP'), ('[de l'éducation.]', 'PP'), ('[moir]', 'NP'), ('[dans chaque f', 'PP'), ('[nous]', 'NP'), ('[demander]', 'PP'), ('[', 'AP'), ('[|]', 'PAUSE'), ('[nan

Mise en pl~ace| au sein de~s différents
l'installation de fumoirs|. En effet, |~
esure a été décidé<e> suite au|~x ~~~~~
tudiants{ fumeurs} qui espéraient pouvo
durant les périodes |plus froides et "au
d'int~empéries. . | C|ette initiative n'

('[pour mieux]', 'PP'), ('[réussir]', 'VN'), ('[par la suite .]', 'PP'), ('[Ces appa-re-ils]', 'NP'), ('[s|~]', 'PP'), ('[|]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('[plus]', 'AdP'), ('[pronon', 'PP'), ('[~fixé]', 'VN'), ('[sur leu{ ~~}', 'PP'), ('[} téléphones]', 'NP'), ('[et]', 'CONJ'), ('[nt]', 'VN'), ('[se sent~~ir |surveillé-s]', 'NP'), ('[ou]', 'CONJ'), ('[|~~empêchés]', 'VN'), ('[P]', 'AP'), ('[perdre]', 'VN'), ('[leur autonomie]', 'NP'), ('[et]', 'CONJ'), ('[leur plaisir]', 'NP')

- How can we find errors automatically ?
- To simplify counts:
 - Reunite curly brackets when needed

('[sur leu{ ~~}', 'PP'), ('[} téléphones]', 'NP') \Rightarrow ('[sur leu{~~}]', 'PP'), ('[téléphones]', 'NP')

Chunking of reconstructed texts

Combining layers: errors and chunks

[('Mise', 'NP'), ('[en place]',
ous]', 'NP'), ('[allons]', 'VN'),
e]', 'PP'), ("de l'éducation.]
moir]', 'NP'), ('[dans chaque f
, ('[nous]', 'NP'), ('[demander]
, 'AP'), ('|', 'PAUSE'), ('[nan

Mise en pl~ace| au sein de~s différents
l'installation de fumoirs|. En effet, |~
esure a été décidé<e> suite au|~x ~~~~~
tudiants{ fumeurs} qui espéraient pouvo
durant les périodes |plus froides et "au
d'int~empéries. . | C|ette initiative n'

('[pour mieux]', 'PP'), ('[réussir]', 'VN'), ('[par la suite .]', 'PP'), ('[Ces appa-re-ils]',
s|~, 'PP'), ('[|]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('[plus]', 'AdP'), ('[pronon
, 'PP'), ('[~fixé]', 'VN'), ('[sur leu{ ~~]', 'PP'), ('[|] téléphones]', 'NP'), ('[et]', 'CONJ'),
nt]', 'VN'), ('[se sent~ir |surveillés]', 'NP'), ('[ou]', 'CONJ'), ('[|~empêchés]', 'VN'), ('
P)', ('[perdre]', 'VN'), ('[leur autonomie]', 'NP'), ('[et]', 'CONJ'), ('[leur plaisir]', 'NP')

- **How can we find errors automatically ?**
- **To simplify counts:**
 - Reunite curly brackets when needed

('[sur leu{ ~~]', 'PP'), ('[|] téléphones]', 'NP') \Rightarrow ('[sur leu{~~}]', 'PP'), ('[téléphones]', 'NP')

- Do the same for other errors (<> for instance)

Chunking of reconstructed texts

Combining layers: errors and chunks

[(['Mise', 'NP'), ('[en place]',
ous]', 'NP'), ('[allons]', 'VN'),
e]', 'PP'), ("[de l'éducation.]
moir]', 'NP'), ('[dans chaque f
, ('[nous]', 'NP'), ('[demander]
, 'AP'), ('|', 'PAUSE'), ('[nan

Mise en pl^ace| au sein de^s différents
l'installation de fumoirs|. En effet, |
esure a été décidé<e> suite au^x ~~~~~
tudiants{ fumeurs} qui espéraient pouvo
durant les périodes |plus froides et "au
d'int^empéries. . | C|ette initiative n'

('[pour mieux]', 'PP'), ('[réussir]', 'VN'), ('[par la suite .]', 'PP'), ('[Ces appa-re-ils]',
s|~.', 'PP'), ('[|]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('[plus]', 'AdP'), ('[pronon
, 'PP'), ('[~fixé]', 'VN'), ('[sur leu{ ~~]', 'PP'), ('[|] téléphones]', 'NP'), ('[et]', 'CONJ'),
nt]', 'VN'), ('[|se sent~~ir |surveillés-]', 'NP'), ('[ou]', 'CONJ'), ('[|~~empêchés]', 'VN'), (P',
, ('[perdre]', 'VN'), ('[leur autonomie]', 'NP'), ('[et]', 'CONJ'), ('[leur plaisir]', 'NP')

- **How can we find errors automatically ?**
- **To simplify counts:**
 - Reunite curly brackets when needed

('[sur leu{ ~~]', 'PP'), ('[|] téléphones]', 'NP') \Rightarrow ('[sur leu{~~}]', 'PP'), ('[téléphones]', 'NP')

- Do the same for other errors (<> for instance)
- Process pauses simultaneously

Chunking of reconstructed texts

Combining layers: errors and chunks

```
[('['Mise]', 'NP'), ('[en place]', 'NP'), ('[ous]', 'NP'), ('[allons]', 'VN'), ('[e]', 'PP'), ("de l'éducation.]", 'PP'), ('[dans chaque f', 'PP'), ('[nous]', 'NP'), ('[demander]', 'PP'), ('[AP]', '|', 'PAUSE'), ('[nan
```

Mise en place| au sein de~s différents
l'installation de fumoirs|. En effet, |
e] mesure a été décidé<e> suite au~x ~~~~~
moir]', 'NP'), ('[dans chaque f', 'PP'), ('[nous]', 'NP'), ('[demander]', 'PP'), ('[AP]', '|', 'PAUSE'), ('[nan



```
('[pour mieux]', 'PP'), ('[réussir]', 'VN'), ('[par la suite .]', 'PP'), ('[Ces appa-re-ils]', 'NP'), ('[s|~', 'PP'), ('[[]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('[plus]', 'AdP'), ('[pronon', 'PP'), ('[~fixé]', 'VN'), ('[sur leu{ ~-}', 'PP'), ('[} téléphones]', 'NP'), ('[et]', 'CONJ'), ('[nt]', 'VN'), ('[se sent~ir |surveillés]', 'NP'), ('[ou]', 'CONJ'), ('[|~empêchés]', 'VN'), ('[P]', 'PP'), ('[perdre]', 'VN'), ('[leur autonomie]', 'NP'), ('[et]', 'CONJ'), ('[leur plaisir]', 'NP')
```

('[pour mieux]', 'PP'), ('[réussir]', 'VN'), ('[par la suite .]', 'PP'), ('[Ces appa-re-ils]', 'NP'), ('[s|~', 'PP'), ('[[]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('[plus]', 'AdP'), ('[pronon', 'PP'), ('[~fixé]', 'VN'), ('[sur leu{ ~-}', 'PP'), ('[} téléphones]', 'NP'), ('[et]', 'CONJ'), ('[nt]', 'VN'), ('[se sent~ir |surveillés]', 'NP'), ('[ou]', 'CONJ'), ('[|~empêchés]', 'VN'), ('[P]', 'PP'), ('[perdre]', 'VN'), ('[leur autonomie]', 'NP'), ('[et]', 'CONJ'), ('[leur plaisir]', 'NP')

- How can we find errors automatically ?
- To simplify counts:
 - Reunite curly brackets when needed

```
('[sur leu{ ~-}', 'PP'), ('[} téléphones]', 'NP') ⇨ ('[sur leu{~-}', 'PP'), ('[téléphones]', 'NP')
```

- Do the same for other errors (<> for instance)
- Process pauses simultaneously

Resulting structure

```
(['sur leu{~-}', 'PP'], ('[téléphones]', 'NP'), ('[aux cours].', 'PP'), ('[|', 'PAUSE'], ('[auront]', 'VN'), ('[plus]', 'AdP'), ('[<-à> voir]', 'PP'), ('[s', 'NP'), ('[|', 'PAUSE'], ('[s<~->', 'CONJ'], ('[appels téléph-oniques]
```

(['sur leu{~-}', 'PP'], ('[téléphones]', 'NP'), ('[aux cours].', 'PP'), ('[|', 'PAUSE'], ('[auront]', 'VN'), ('[plus]', 'AdP'), ('[<-à> voir]', 'PP'), ('[s', 'NP'), ('[|', 'PAUSE'], ('[s<~->', 'CONJ'], ('[appels téléph-oniques]

- Pause in-between chunks
- Pause within chunks
- Insertion of one character (following 2 deletions)
- Insertion of a string of characters (that was then deleted)

Chunking of reconstructed texts

Combining layers: errors and chunks

```
[('[Mise]', 'NP'), ('[en place]', 'NP'), ('[ous]', 'NP'), ('[allons]', 'VN'), ('[e]', 'PP'), ("de l'éducation.]", 'PP'), ('[dans chaque f', 'PP'), ('[nous]', 'NP'), ('[demander]', 'PP'), ('[AP]', '|', 'PAUSE'), ('[nan
```

Mise en place] au sein de ns différents
l'installation de fumoirs|. En effet, |
e] mesure a été décidé<e> suite au~x ~~~~~
moir]', 'NP'), ('[dans chaque f', 'PP'), ('[studants{ fumeurs} qui espéraient pouvo
, ('[nous]', 'NP'), ('[demander]', 'PP'), ('[durant les périodes |plus froides et "au
, 'AP'), ('|', 'PAUSE'), ('[nan d'int~empéries. . | C]ette initiative n'



```
('[pour mieux]', 'PP'), ('[réussir]', 'VN'), ('[par la suite .]', 'PP'), ('[Ces appa-re-ils]', 'NP'), ('[s|~', 'PP'), ('[|]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('[plus]', 'AdP'), ('[pronon', 'PP'), ('[~fixé]', 'VN'), ('[sur leu{ ~-}', 'PP'), ('[|] téléphones]', 'NP'), ('[et]', 'CONJ'), ('[nt]', 'VN'), ('[|se sent~ir |surveillés]', 'NP'), ('[ou]', 'CONJ'), ('[|~empêchés]', 'VN'), ('[P]', 'PP'), ('[perdre]', 'VN'), ('[leur autonomie]', 'NP'), ('[et]', 'CONJ'), ('[leur plaisir]', 'NP')
```

- How can we find errors automatically ?
- To simplify counts:
 - Reunite curly brackets when needed

```
('[sur leu{ ~-}', 'PP'), ('[|] téléphones]', 'NP') ⇨ ('[sur leu{~-}', 'PP'), ('[téléphones]', 'NP')
```

- Do the same for other errors (<> for instance)
- Process pauses simultaneously

Resulting structure

```
(['sur leu{~-}', 'PP'], ('[téléphones]', 'NP'), ('[aux cours|.]', 'PP'), ('|', 'PAUSE'), ('[auront]', 'VN'), ('[plus]', 'AdP'), ('[<-à> voir]', 'PP'), ('[s]', 'NP'), ('|', 'PAUSE'), ('[s<~->', 'PP'), ('[ils]', 'NP'), ('[CONJ]', 'PP'), ('[appels téléph-oniques]', 'NP')
```

Pause in-between chunks
Pause within chunks
Insertion of one character (following 2 deletions)
Insertion of a string of characters (that was then deleted)

- Simplify counting of all types
- Retrieve context of all types

Chunking of reconstructed texts

Combining layers: errors and chunks

[[('Mise', 'NP'), ('en place', 'NP'), ('nous', 'NP'), ('allons', 'VN'), ('e', 'PP'), ("de l'éducation.", 'PP'), ('moir', 'NP'), ('dans chaque f', 'PP'), ('nous', 'NP'), ('demander', 'PP'), ('AP'), ('|', 'PAUSE'), ('nan

Mise en place au sein de différents lieux de l'installation de fumeurs. En effet, la mesure a été décidée suite au vote des étudiants fumeurs qui espéraient pouvoir fumer pendant les périodes plus froides et au cours de l'hiver. Cette initiative n'



('pour mieux', 'PP'), ('réussir', 'VN'), ('par la suite .', 'PP'), ('[Ces appareils]', 'PP'), ('s', 'PP'), ('[', 'NP'), ('par une attention', 'PP'), ('plus', 'AdP'), ('pronon', 'PP'), ('[~fixé]', 'VN'), ('sur leu{ ~}', 'PP'), ('[] téléphones', 'NP'), ('et', 'CONJ'), ('nt', 'VN'), ('[se sent~ir |surveillés]', 'NP'), ('ou', 'CONJ'), ('[~empêchés]', 'VN'), ('P', 'PP'), ('perdre', 'VN'), ('leur autonomie]', 'NP'), ('et', 'CONJ'), ('leur plaisir', 'NP')]

- How can we find errors automatically ?
- To simplify counts:
 - Reunite curly brackets when needed

('sur leu{ ~}', 'PP'), ('[] téléphones', 'NP') ⇨ ('sur leu{~}', 'PP'), ('[téléphones]', 'NP')

- Do the same for other errors (<> for instance)
- Process pauses simultaneously

Resulting structure

(['sur leu{~}', 'PP'], ('[téléphones]', 'NP'), ('aux cours.', 'PP'), ('|', 'PAUSE'), ('auront', 'VN'), ('plus', 'AdP'), ('<-à voir', 'PP'), ('s', 'PP'), ('|', 'PAUSE'), ('s<~>', 'PP'), ('ils', 'NP'), ('appels téléph-oniques', 'NP')]

Pause in-between chunks
 Pause within chunks
 Insertion of one character (following 2 deletions)
 Insertion of a string of characters (that was then deleted)

- Simplify counting of all types
- Retrieve context of all types

qui NP

Left context
 chunk + type

<n> NP

Character insertion

auront VN

Right context
 chunk + type

Chunking of reconstructed texts

Combining layers: errors and chunks

[[['Mise', 'NP'), ('[en place]', 'NP'), ('[ous]', 'NP'), ('[allons]', 'VN'), ('[e]', 'PP'), ('[de l'éducation.]', 'PP'), ('[moir]', 'NP'), ('[dans chaque f', 'PP'), ('[nous]', 'NP'), ('[demander]', 'PP'), ('[', 'AP'), ('[|', 'PAUSE'), ('[nan

Mise en place| au sein de|s différents| l'installation de fumoirs|. En effet, |e| sure a été décidé<e> suite au|~x ~~~~~| étudiants{ fumeurs} qui espéraient pouvo| durant les périodes |plus froides et "au| d'int~empéries. . | C|ette initiative n'



('[pour mieux]', 'PP'), ('[réussir]', 'VN'), ('[par la suite .]', 'PP'), ('[Ces appa-re-ils]', 'NP'), ('[s|~', 'PP'), ('[|]', 'NP'), ('[~~~~~par |une attention]', 'PP'), ('[plus]', 'AdP'), ('[pronon', 'PP'), ('[~fixé]', 'VN'), ('[sur leu{ ~~}', 'PP'), ('[|} téléphones]', 'NP'), ('[et]', 'CONJ'), ('[nt]', 'VN'), ('[|se sent~ir |surveillé-s]', 'NP'), ('[ou]', 'CONJ'), ('[|~empêchés]', 'VN'), ('[P]', 'PP'), ('[perdre]', 'VN'), ('[leur autonomie]', 'NP'), ('[et]', 'CONJ'), ('[leur plaisir]', 'NP')

- How can we find errors automatically ?
- To simplify counts:

- Reunite curly brackets when needed

('[sur leu{ ~~}', 'PP'), ('[|} téléphones]', 'NP') ⇨ ('[sur leu{~~}]', 'PP'), ('[téléphones]', 'NP')

- Do the same for other errors (<> for instance)
- Process pauses simultaneously

Resulting structure

(['sur leu{~~}'], 'PP'), ('[téléphones]', 'NP')
 ('[aux cours|.]', 'PP'), ('[|', 'PAUSE'), ('[auront]', 'VN'), ('[plus]', 'AdP'), ('[<-à> voir]', 'PP'), ('[s]', 'NP'), ('[|', 'PAUSE'), ('[s<~~>]', 'PP'), ('[ils]', 'NP'), ('[appels téléph-oniques]', 'NP')

Pause in-between chunks
 Pause within chunks
 Insertion of one character (following 2 deletions)
 Insertion of a string of characters (that was then deleted)

- Simplify counting of all types
- Retrieve context of all types

qui NP

Left context
 chunk + type

<n> NP

Character insertion

auront VN

Right context
 chunk + type

PAUSE

Left context
 chunk + type

||es ~~~~~ professeurs NP

Characters deletion

confisquent VN

Right context
 chunk + type

Chunking of reconstructed texts

Preliminary quantitative analysis

```
(('[sont basé]', 'VN'), ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(('[', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
(('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
('[en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
(('[éno~rmément]', 'AdP'), ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
(('[malgré tout]', 'PP'), ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
(('[de cette médecine]', 'PP'), ('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
(('[la médecine alternative]', 'NP'), ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
(('[médi]', 'NP'), ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
(('[peut]', 'VN'), ('[malgré tout]', 'PP'), ('[pos<~é>]', 'VN'))
(('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
('[sont basé]', 'VN'), ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(('[', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
(('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(('[en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
(('[éno~rmément]', 'AdP'), ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
(('[malgré tout]', 'PP'), ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
(('[de cette médecine]', 'PP'), ('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
(('[la médecine alternative]', 'NP'), ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|ecine]', 'PP'))
(['', 'AdP'], ('[certain-s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant]', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
(['éno-rmément]', 'AdP'], ('[mi<-t>]', 'VN'), ('[en avant]', 'PP'))
(['[malgré tout]', 'PP'], ('[pos<-é>]', 'VN'), ('[beaucoup]', 'AdP'))
(['[de cette médecine]', 'PP'], ('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
(['[la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
(['[médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
(['[peut]', 'VN'], ('[malgré tout]', 'PP'), ('[pos<-é>]', 'VN'))
(['<s>e basant]', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|ecine]', 'PP'))
(['', 'AdP'], ('[certain-s remède<s>]', 'NP'), ('[sont]', 'VN'))
(['<s>e basant]', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
(['éno-rmément]', 'AdP'], ('[mi<-t>]', 'VN'), ('[en avant]', 'PP'))
(['[malgré tout]', 'PP'], ('[pos<-é>]', 'VN'), ('[beaucoup]', 'AdP'))
(['[de cette médecine]', 'PP'], ('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
(['[la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(([, 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
((['médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
((['peut]', 'VN'], ('[malgré tout]', 'PP'], ('[pos<~é>]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(([, 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(([, 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
((['médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
((['peut]', 'VN'], ('[malgré tout]', 'PP'], ('[pos<~é>]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(([, 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('[<s>e basant]', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors
- Measure relations between/among those segmentations

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(([, 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
((['médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
((['peut]', 'VN'], ('[malgré tout]', 'PP'], ('[pos<~é>]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
([, 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
(['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors
- Measure relations between/among those segmentations

General hypotheses:

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(['.', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
((['médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
((['peut]', 'VN'], ('[malgré tout]', 'PP'], ('[pos<~é>]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(['.', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors
- Measure relations between/among those segmentations

General hypotheses:

- **Pausal segmentation** tends to correspond to **chunk segmentation half of the time: on 3034 pauses, 1480 occur in-between two chunks (48,78 %)**

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
((['', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('<s>e basant', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
((['médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
((['peut]', 'VN'], ('[malgré tout]', 'PP'], ('[pos<~é>]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(|', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('<s>e basant', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors
- Measure relations between/among those segmentations

General hypotheses:

- **Pausal segmentation** tends to correspond to **chunk segmentation half of the time: on 3034 pauses, 1480 occur in-between two chunks (48,78 %)**
- **Pauses** seem to occur **mostly before: noun chunks (610 noun chunks following pauses, 17% of noun chunks follow a pause)**

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(['', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
(['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
(['[malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
(['[de cette médecine]', 'PP'], ('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,
(['|', 'PAUSE'], ('[elle]', 'NP'), ('|', 'PAUSE'))
(['[la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
(['[médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(['|', 'PAUSE'], ('[au final]', 'PP'), ('|', 'PAUSE'))
(['[peut]', 'VN'], ('[malgré tout]', 'PP'), ('[pos<~é>]', 'VN'))
(['[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(['', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
(['[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
(['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
(['[malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
(['[de cette médecine]', 'PP'], ('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,
(['|', 'PAUSE'], ('[elle]', 'NP'), ('|', 'PAUSE'))
(['[la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors
- Measure relations between/among those segmentations

General hypotheses:

- **Pausal segmentation** tends to correspond to **chunk segmentation half of the time: on 3034 pauses, 1480 occur in-between two chunks (48,78 %)**
- **Pauses** seem to occur **mostly before: noun chunks (610 noun chunks following pauses, 17% of noun chunks follow a pause)**
- **Pauses** seem to occur **mostly after: preposition chunks (539 preposition chunks preceding pauses, 15,5% of pp chunks precede a pause) and adjectival chunks (32 adjectival chunks preceding pauses, 16% of ap chunks precede a pause)**

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
((['', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('<s>e basant', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
((['médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
((['peut]', 'VN'], ('[malgré tout]', 'PP'], ('[pos<~é>]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi|~ecine]', 'PP'))
(|', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('<s>e basant', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors
- Measure relations between/among those segmentations

General hypotheses:

- **Pausal segmentation** tends to correspond to **chunk segmentation half of the time: on 3034 pauses, 1480 occur in-between two chunks (48,78 %)**
- **Pauses** seem to occur **mostly before: noun chunks (610 noun chunks following pauses, 17% of noun chunks follow a pause)**
- **Pauses** seem to occur **mostly after: preposition chunks (539 preposition chunks preceding pauses, 15,5% of pp chunks precede a pause) and adjectival chunks (32 adjectival chunks preceding pauses, 16% of ap chunks precede a pause)**

We could go further...

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi-ecine]', 'PP'))
(('[', 'AdP'], ('[certain-s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
(('[éno-rmément]', 'AdP'], ('[mi<-t>]', 'VN'), ('[en avant]', 'PP'))
(('[malgré tout]', 'PP'), ('[pos<-é>]', 'VN'), ('[beaucoup]', 'AdP'))
(('[de cette médecine]', 'PP'), ('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
(('[la médecine alternative]', 'NP'), ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
(('[médi]', 'NP'), ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
(('[peut]', 'VN'), ('[malgré tout]', 'PP'), ('[pos<-é>]', 'VN'))
(('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi-ecine]', 'PP'))
(('[', 'AdP'], ('[certain-s remède<s>]', 'NP'), ('[sont]', 'VN'))
(('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
(('[éno-rmément]', 'AdP'], ('[mi<-t>]', 'VN'), ('[en avant]', 'PP'))
(('[malgré tout]', 'PP'), ('[pos<-é>]', 'VN'), ('[beaucoup]', 'AdP'))
(('[de cette médecine]', 'PP'), ('[<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
(('[la médecine alternative]', 'NP'), ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors
- Measure relations between/among those segmentations

General hypotheses:

- **Pausal segmentation** tends to correspond to **chunk segmentation half of the time: on 3034 pauses, 1480 occur in-between two chunks (48,78 %)**
- **Pauses** seem to occur **mostly before: noun chunks (610 noun chunks following pauses, 17% of noun chunks follow a pause)**
- **Pauses** seem to occur **mostly after: preposition chunks (539 preposition chunks preceding pauses, 15,5% of pp chunks precede a pause) and adjectival chunks (32 adjectival chunks preceding pauses, 16% of ap chunks precede a pause)**

We could go further...

- What **type of chunks** contains the **most errors?**

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi~ecine]', 'PP'))
((['', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('<s>e basant', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
((['médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
((['peut]', 'VN'], ('[malgré tout]', 'PP'), ('[pos<~é>]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi~ecine]', 'PP'))
(|', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<~t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<~é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('<s>e basant', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors
- Measure relations between/among those segmentations

General hypotheses:

- **Pausal segmentation** tends to correspond to **chunk segmentation half of the time**: *on 3034 pauses, 1480 occur in-between two chunks (48,78 %)*
- **Pauses** seem to occur **mostly before**: *noun chunks (610 noun chunks following pauses, 17% of noun chunks follow a pause)*
- **Pauses** seem to occur **mostly after**: *preposition chunks (539 preposition chunks preceding pauses, 15,5% of pp chunks precede a pause) and adjectival chunks (32 adjectival chunks preceding pauses, 16% of ap chunks precede a pause)*

We could go further...

- What **type of chunks** contains the **most errors**?
- Do **character insertions** happen most commonly after pauses?

Chunking of reconstructed texts

Preliminary quantitative analysis

```
((['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi~ecine]', 'PP'))
((['', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('<s>e basant', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
((['médi]', 'NP'], ('[cal]', 'PP'), ('|', 'PAUSE'))
(|', 'PAUSE'), ('[au final]', 'PP'), ('|', 'PAUSE'))
((['peut]', 'VN'], ('[malgré tout]', 'PP'), ('[pos<é>]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
(['sont basé', 'VN'], ('[sur les fond<s>]', 'PP'), ('[de cette médi~ecine]', 'PP'))
(|', 'AdP'], ('[certain~s remède<s>]', 'NP'), ('[sont]', 'VN'))
((['<s>e basant', 'PP'], ('[sur les] ancienne<s> croyances,]', 'PP'), ('[les erreurs médic
([en~ effet fort efficace]', 'PP'), ('[sur certain<s> points~.]', 'PP'), ('|', 'PAUSE'))
((['éno~rmément', 'AdP'], ('[mi<t>]', 'VN'), ('[en avant]', 'PP'))
((['malgré tout]', 'PP'], ('[pos<é>]', 'VN'), ('[beaucoup]', 'AdP'))
((['de cette médecine]', 'PP'], ('<s>e basant', 'PP'), ('[sur les] ancienne<s> croyances,
(|', 'PAUSE'), ('[elle]', 'NP'), ('|', 'PAUSE'))
((['la médecine alternative]', 'NP'], ('[peut]', 'VN'), ('[malgré tout]', 'PP'))
```

- Apply these functions to the whole corpus
- Obtain numbers regarding pauses, chunks and errors
- Measure relations between/among those segmentations

General hypotheses:

- **Pausal segmentation** tends to correspond to **chunk segmentation half of the time: on 3034 pauses, 1480 occur in-between two chunks (48,78 %)**
- **Pauses** seem to occur **mostly before: noun chunks (610 noun chunks following pauses, 17% of noun chunks follow a pause)**
- **Pauses** seem to occur **mostly after: preposition chunks (539 preposition chunks preceding pauses, 15,5% of pp chunks precede a pause) and adjectival chunks (32 adjectival chunks preceding pauses, 16% of ap chunks precede a pause)**

We could go further...

- What **type of chunks** contains the **most errors?**
- Do **character insertions** happen most commonly **after pauses?**
- In **what type of chunks** do people generally decide to **revise a preceding portion of text?**

Conclusion

- **Chunks seem to be relevant units for real-time recorded writing**
- **It is possible to adapt existing automatic tools to real-time recorded writing**
- **Some hypotheses regarding the overlap of different segmentations seem to be confirmed (Half of pauses occur in-between two chunks)**
- **Similar studies must be led for errors and their relations to chunks and pauses**
- **We need to focus on a better chunking performance and a stricter processing of our data for further analysis**

Annexes

Data

- 3 csv files containing information on the participants' written productions

| ID | charge | n_burst | categ | burst | charBurst | startPos | endPos | docLength | debut_burst | duree_burst | duree_pause | duree_cycle | longueur_burst |
|------|--------|---------|-------|---|--|----------|--------|-----------|-------------|-------------|-------------|-------------|----------------|
| P+S1 | + | 1 | P | Dans le cadre de la médecine traditionnelle, | Dans_le_cadre_de_la_médecine_traditionnelle, | 0 | 45 | 45 | 22,397 | 10,464 | | 10,464 | 45 |
| P+S1 | + | 2 | P | nous voici face à un dilemme. | nous_voici_face_à_un_dilemme, | 45 | 76 | 76 | 38,557 | 7,296 | 5,696 | 12,992 | 31 |
| P+S1 | + | 3 | P | Beaucoup critiqué cette dernière s'avère parfois en effet fort efficace sur certaines points. | Beaucoup_critiqué_cette_dernière_s'avère_parfois_ene_effet_fort_efficace_sur_certaines_points, | 76 | 170 | 170 | 54,837 | 21,672 | 8,984 | 30,656 | 94 |
| P+S1 | + | 4 | P | En effet si l'ont y réfléchié plus longuement beaucoup de nos méthodes actuelles se sont basé sur les fond de cette médi | En_effet_si_l'ont_y_réfléchié_plus_longuement_beaucoup_de_nos_méthodes_actuelles_se_sont_basé_sur_les_fond_de_cette_médi | 170 | 290 | 290 | 85,301 | 21,12 | 8,792 | 29,912 | 120 |
| P+S1 | + | 5 | RB | ecine alternative | ecine_alternative | 290 | 306 | 306 | 108,453 | 3,584 | 2,032 | 5,616 | 17 |
| P+S1 | + | 6 | P | , certains remède sont encore utilisés | ,_certains_remède_sont_encore_utilisés, | 306 | 345 | 345 | 116,061 | 10,767 | 4,024 | 14,791 | 39 |
| P+S1 | + | 7 | P | à nos jours | à_nos_jours | 345 | 356 | 356 | 129,972 | 2,512 | 3,144 | 5,656 | 11 |
| P+S1 | + | 8 | P | . | . | 356 | 358 | 358 | 134,38 | 3,464 | 1,896 | 5,36 | 2 |
| P+S1 | + | 9 | R | s | s | 323 | 347 | 359 | 145,38 | 96 | 7,536 | 7,632 | 1 |
| P+S1 | + | 10 | RB | | | 358 | 359 | 360 | 160,22 | 0,24 | 14,744 | 14,984 | 1 |
| P+S1 | + | 11 | R | s | s | 276 | 361 | 361 | 167,164 | 0,12 | 6,704 | 6,824 | 1 |
| P+S1 | + | 12 | P | De plus nous pouvons souligner le fait que la médecine alternative | De_plus_nous_pouvons_souligner_le_fait_que_la_médecine_alternative, | 363 | 428 | 428 | 184,604 | 18,136 | 17,32 | 35,456 | 67 |
| P+S1 | + | 13 | RB | n'utilise pas que les croyances | n'utilisepas_que_les_croyances | 428 | 459 | 459 | 204,732 | 7,736 | 1,992 | 9,728 | 31 |
| P+S1 | + | 14 | P | , | , | 459 | 461 | 461 | 214,996 | 0,4 | 2,528 | 2,928 | 2 |
| P+S1 | + | 15 | P | les remèdes naturelles sont en effet | les_remèdes_naturelles_sont_en_effet | 461 | 483 | 497 | 218,02 | 6,183 | 2,624 | 8,807 | 36 |
| P+S1 | + | 16 | RB | | | 482 | 483 | 495 | 228,123 | 248 | 3,92 | 4,168 | 0 |
| P+S1 | + | 17 | P | t énormément mis en avant contrairement à la médecine actuelle qui même si elle s'avère efficace utilise pour la plupart du temps des médicaments plus ou moins chimique pouvant parfois être plus nocif que la maladie en elle-même. | t_beaucoup_mis_en_avant_contrairement_à_la_médecine_actuelle_qui_même_si_elle_s'avère_efficace_utilise_pour_la_plupart_du_temps_des_médicaments_plus_ou_moins_chimiqu_pouvant_parfois_être_plus_nocif_que_la_maladie_en_elle-même, | 495 | 724 | 724 | 232,035 | 64,68 | 3,664 | 68,344 | 230 |

How to use data to rebuild the texts?

For each burst of each participant:

— Add the burst at the *starting position of the cursor*

— If it contains **deletion characters** (“⊠” or “⊡”):

charBurst

— Replace the deleted characters with squiggly lines (“~”)

— If it is **added** to the existing text:

startPos

docLength

— Separate it from the existing text with a pipe (“|”)

— If it is **inserted** within the existing text:

startPos

docLength

— If it is a single character:

charBurst

— Enclose the burst in chevrons (“< >”)

— If it is a sequence of characters:

charBurst

— Enclose the burst in curly brackets (“{ }”)

That ⊠ ⊠ is _ is _ an _ exe ⊠ ample

Th ~ is _ is _ an _ ex ~ ample

Th ~ is _ is _ an _ ex ~ ample | _ of _ a _ text.

Example < s > _ of _ insertion

Insertion _ of { _ multiple } _ characters

GitHub repository for the reconstruction scripts:

https://github.com/ims510/pos-tagger/tree/aj_stage