

# Annotation de corpus avec Gate



Alexane Jouglar

# Déroulement de mon projet

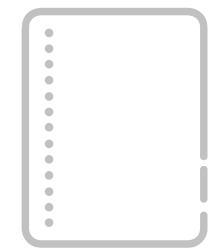
---

## Aperçu



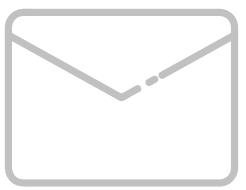
### Initialisation du projet

Sélection des livres et  
définition du corpus.



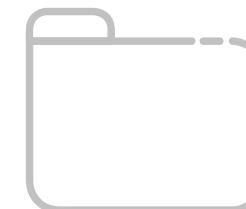
### Schémas d'annotation

Création des schémas  
d'annotation :  
explications et  
illustration.



### Dictionnaires (non-)flexibles

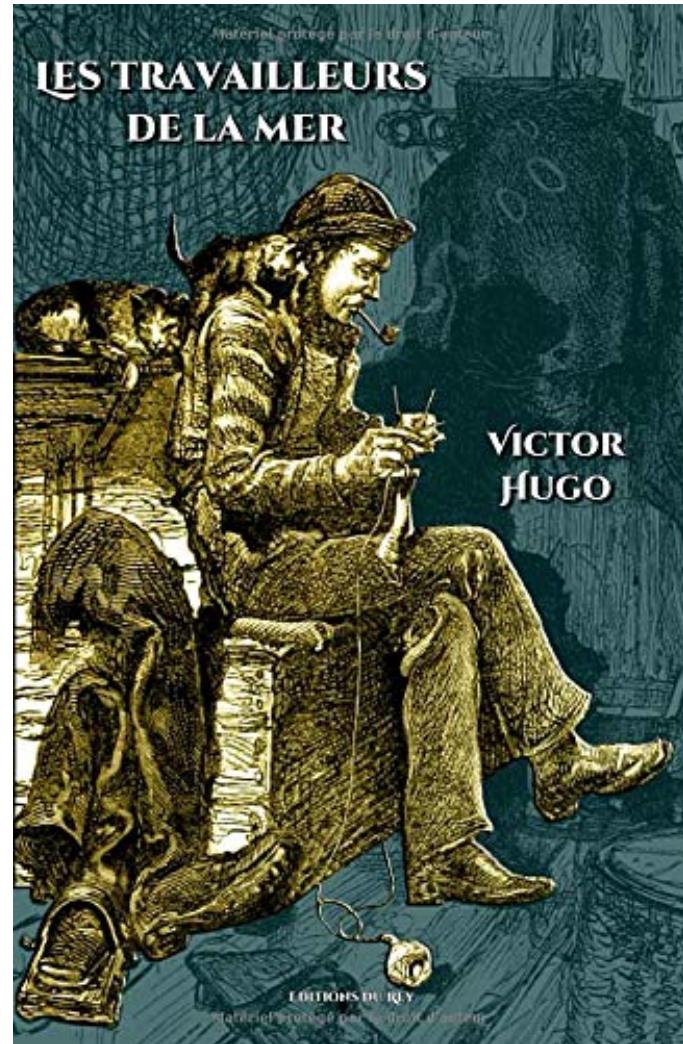
Création des  
dictionnaires : flexibles  
et non-flexibles (en  
fonction du corpus).



### Grammaires JAPE

Création des  
grammaires JAPE :  
explications et  
illustration.

# Initialisation : sélection des livres et définition du corpus



Deux livres :

- *Alice's adventures in Wonderland*, Lewis Caroll, 1865
- *Les Travailleurs de la mer*, Victor Hugo, 1866

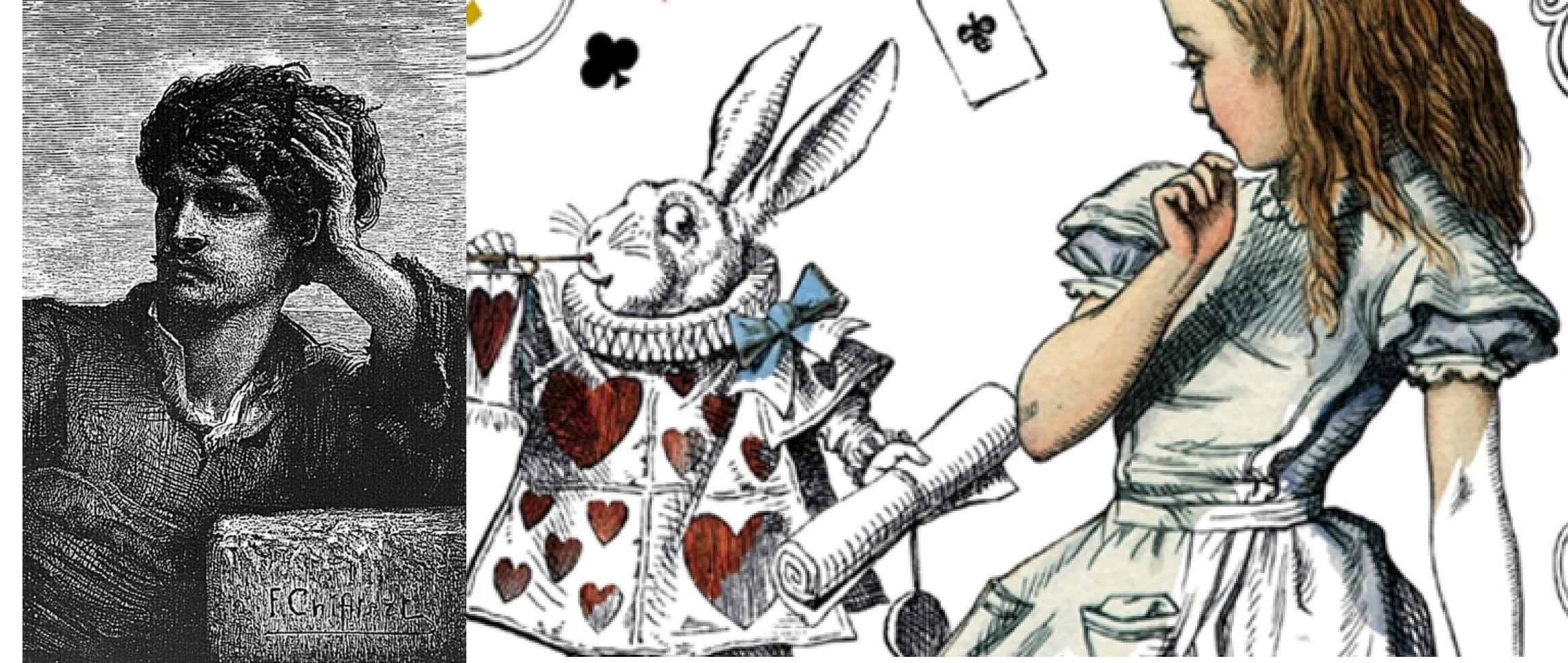


# Schémas d'annotation

## Annotations manuelles

Cinq  
schémas  
d'annotation

---



- **Structure** : titre et dialogue
- **Personnage** : principal et secondaire
- **Pensée** : à voix haute ou sans guillemets
- **Lieux** : intérieurs ou extérieurs
- **Sexe** : Figurant féminin ou Figurant masculin (schéma utilisé uniquement pour le corpus français)

# Schémas d'annotation (2)

<input checked="" type="checkbox"/>	lieux
<input checked="" type="checkbox"/>	pensées
<input checked="" type="checkbox"/>	personnages
<input checked="" type="checkbox"/>	structure

{personnages=Personnage principal}  
{personnages=Personnage principal, safe.regex=true}  
{lieux=lieu extérieur}  
{pensées=pensées à voix haute (guillemets)}  
{personnages=Personnage principal, safe.regex=true}

## CHAPTER I

### Down the Rabbit-Hole

Alice was beginning to get very tired of sitting by her sister on the bank, and of having nothing to do: once or twice she had peeped into the book her sister was reading, but it had no pictures or conversations in it, 'and what is the use of a book,' thought Alice, 'without pictures or conversation?'

So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.

There was nothing so very remarkable in that; nor did Alice think it so very much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at this, but at the time it all seemed quite natural); but when the Rabbit actually took a watch out of its waistcoat - pocket , and looked at it, and then hurried on, Alice started to her feet, for it flashed across her mind that she had never before seen a rabbit with either a waistcoat-pocket, or a watch to take out of it, and burning with curiosity, she ran across the field after it, and fortunately was just in time to see it pop down a large rabbit-hole under the hedge.

UN MOT ÉCRIT SUR UNE PAGE BLANCHE

La Christmas de 182. fut remarquable à Guernesey. Il neigea ce jour-là.  
mémorable, et la neige fait événement.

Le matin de cette Christmas, la route qui longe la mer de Saint-Pierre-Po  
l'aube. Vers neuf heures, peu après le lever du soleil, comme ce n'était p  
e la première tour de la seconde tour, il n'y avait que trois passants, un enfant, un  
épisode. L'homme, jeune encore, semblait quelque chose comme un ouvrier ou un  
usage de gros drap brun, et un pantalon à jambières goudronnées, ce qui paraissait  
ille. Ses épais souliers de cuir brut, aux semelles garnies de gros clous, laissaient sur  
plus délicate des transitions, l'adolescence, les deux crépuscules mêlés, le  
. L'homme ne la remarquait pas.

## *Les Travailleurs de la mer : 4 majorType*

- **Lieux** : pays, îles
- **Description physique** : grandeur, forme, beauté/âge,  
cheveux, force
- **Références temporelles** : passé, général
- **Matières** : tissus, couleurs

# Dictionnaires non flexibles

Hash gazetteer  
en français



down, down. There was nothing else to do, so Alice soon began talking again. 'Dinah was the cat.) 'I hope they'll remember her saucer of milk at tea-time. Dinah are no mice in the air, I'm afraid, but you might catch a bat, and that's very like a n r? And here Alice began to get rather sleepy, and went on saying to herself, in a d at bats?' and sometimes, 'Do bats eat cats?' for, you see, as she couldn't answer e he put it. She felt that she was dozing off, and had just begun to dream that she w to her very earnestly, 'Now, Dinah, tell me the truth: did you ever eat a bat?' wh a heap of sticks and dry leaves, and the fall was over.

## Alice's adventures in wonderland : 3

---

### majorType

MajorType issu d'un **dictionnaire non-flexible** :

- **Description physique** : couleur, taille, apparence (liste d'adjectifs)
- **Espèce** : petits mammifères, insectes

MajorType issu d'un **dictionnaire flexible** :

- **Verbes** : de mouvement, de parole, d'expression

Suddenly she came upon a little three-legged table, all made of solid glass; there was nothing on it except a tiny golden key, and Alice's first thought was that it might belong to one of the doors of the hall; but, alas! either the locks were too large, or the key was too small, but at any rate it would not open any of them. However, on the second time round, she came upon a low curtain she had not noticed before, and behind it was a little door about fifteen inches high: she tried the little golden key in the lock, and to her great

# Dictionnaires flexibles et non flexibles

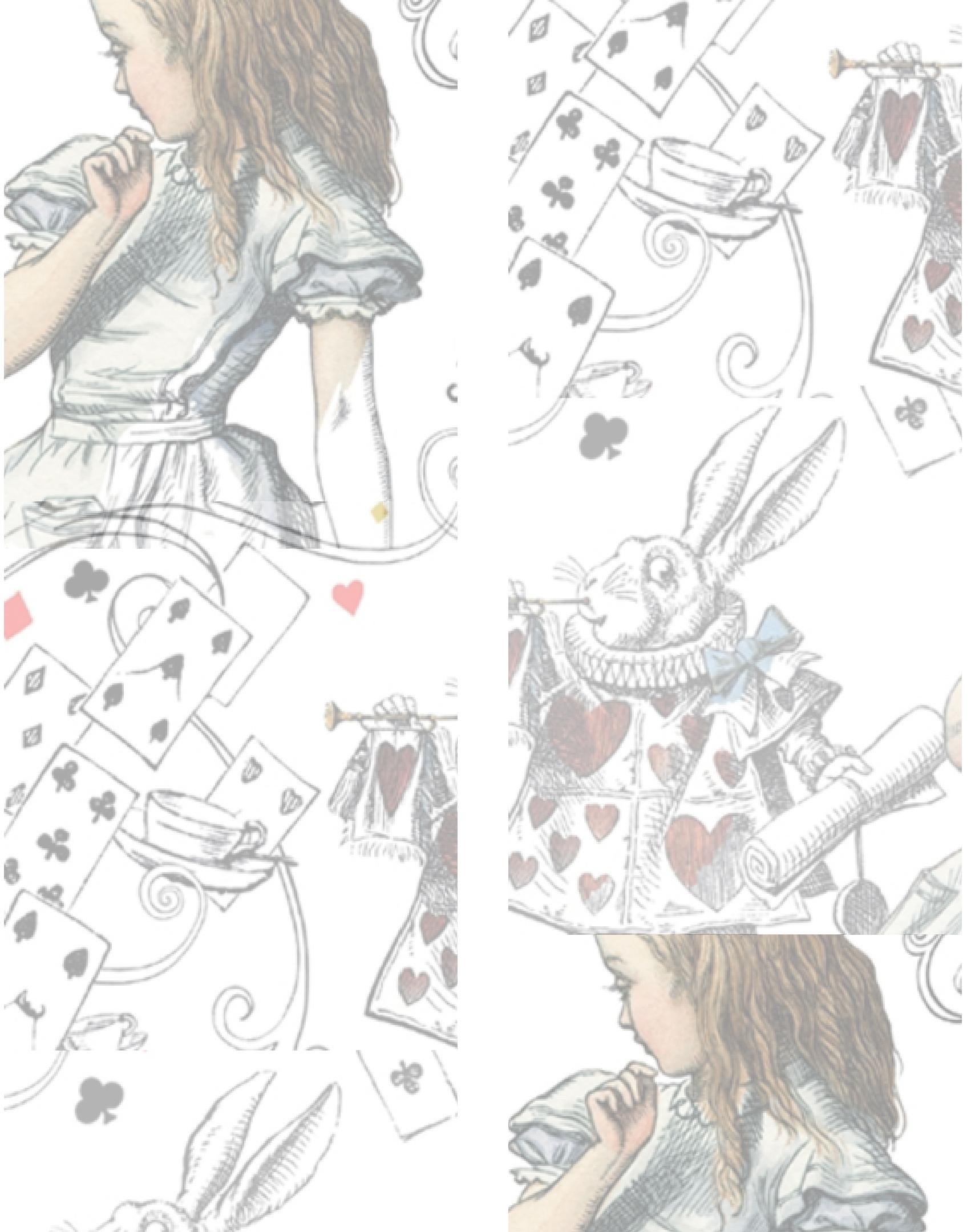
Hash gazetteer et flexible gazetteer en anglais

# Ma grammaire JAPE

Application au corpus anglais

## 4 annotations

- 1 première appelée "Moment" qui correspond à ma première grammaire JAPE
- Les trois autres sont des sous-ensembles de celle-ci (diapo suivante)



1.

So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.

2.

So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.

3.

So she was considering in her own mind (as well as she could, for the hot day made her feel very sleepy and stupid), whether the pleasure of making a daisy-chain would be worth the trouble of getting up and picking the daisies, when suddenly a White Rabbit with pink eyes ran close by her.

## 1. BeforeAdverb

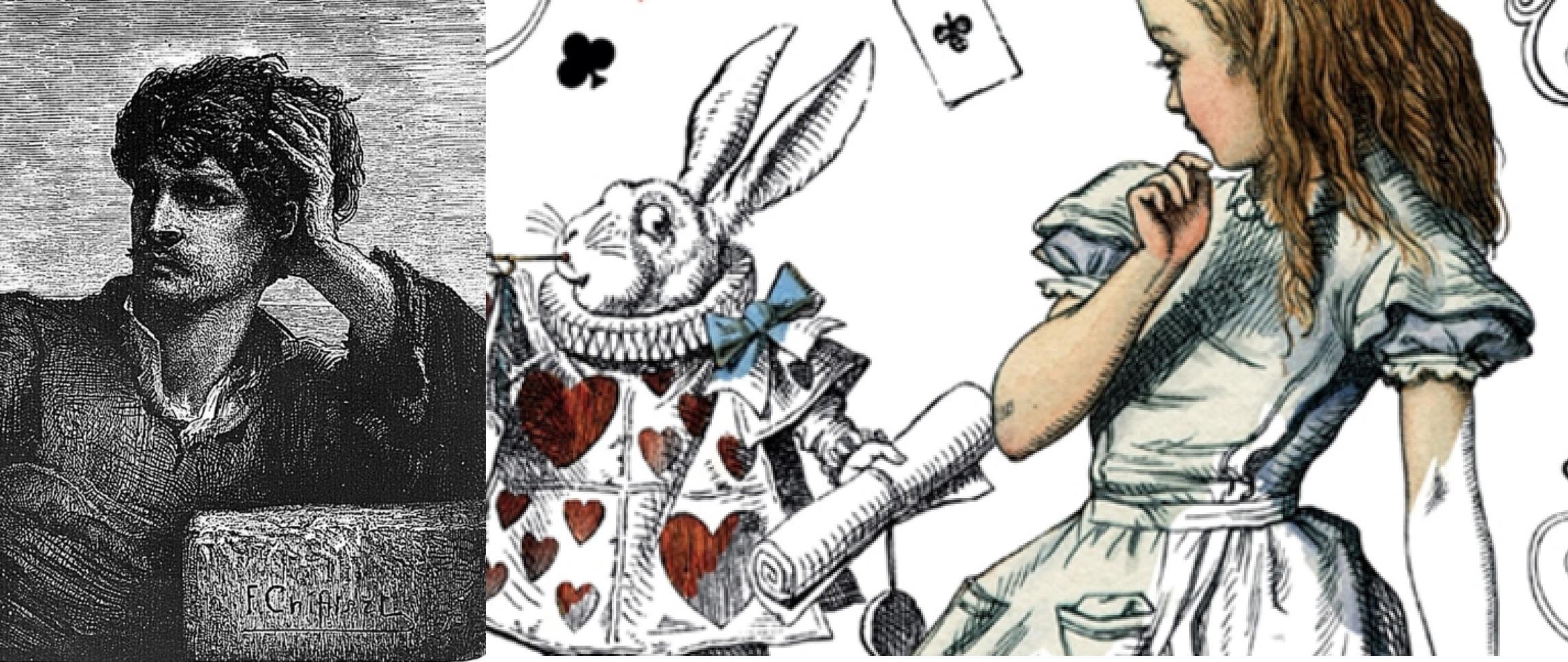
Ce que l'on trouve avant le token qui introduit l'action

## 2. Word

Dans l'exemple présenté ci-dessus : "suddenly". Ce type de mot introduit une action qui a une importance dans l'histoire.

## 3. Action

Ce qui vient après le token "suddenly". C'est généralement un moment important de l'histoire.



## Deux traitements conditionnels

---

# Traitements conditionnels

Sans compter l'application de la reconnaissance de langue (TextCat Language Identification), **deux séries de traitements conditionnels** ont vu le jour : **une pour chaque corpus** (les traitements pour le corpus anglais étant plus nombreux bien entendu).

# Mes deux traitements conditionnels

---

 ANNIE English Tokeniser	ANNIE English Tokeniser
 ANNIE Sentence Splitter	ANNIE Sentence Splitter
 ANNIE POS Tagger	ANNIE POS Tagger
 Analyse_morphologique	GATE Morphological analyser
 Dico_non_flexible_en	Hash Gazetteer
 Mon_dictionnaire_flexible_en	Flexible Gazetteer

## Traitements pour mon corpus anglais

Je n'y mets pas ici ma grammaire JAPE que j'ai appliquée en amont de mon travail et avant toute autre annotation. La raison : GATE n'appliquait ma JAPE que si j'y ajoutais en premier lieu un Document Reset.

 ANNIE English Tokeniser	ANNIE English Tokeniser
 ANNIE Sentence Splitter	ANNIE Sentence Splitter
 Hash Gazetteer 000E2	Hash Gazetteer

## Traitements pour corpus français

Peu de traitements, parce que pas de TreeTagger french installé. Le hash Gazetteer qui n'a pas été renommé contient le dictionnaire non-flexible français.

# Conclusion...

## Avantages et inconvénients de l'interface GATE

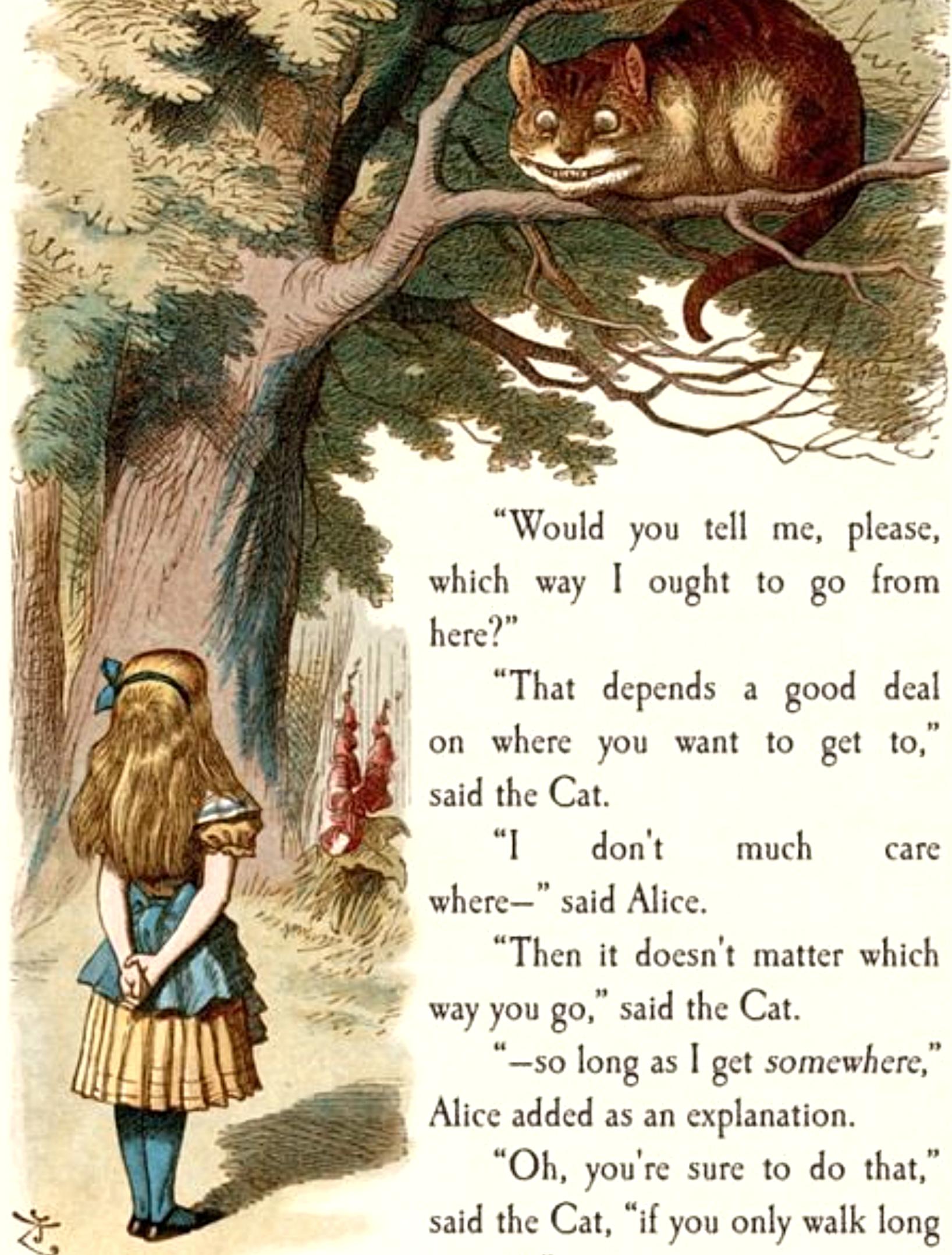
---

## Avantages et inconvénients de GATE

---

- Prise en main rapide et intuitive
- L'interface est assez vite limitée parce qu'elle ne supporte pas la complexité.
- Il faut en amont réfléchir à tous les traitements voulus avant de créer les annotations (les groupes d'annotation ne peuvent être modifiés)
- Gate Developer offre une infinité de traitements possibles.
- Certains plugins qui n'ont pas été utilisé dans le projet permettent d'appliquer des traitements à n'importe quelle langue.





"Would you tell me, please,  
which way I ought to go from  
here?"

"That depends a good deal  
on where you want to get to,"  
said the Cat.

"I don't much care  
where—" said Alice.

"Then it doesn't matter which  
way you go," said the Cat.

"—so long as I get somewhere,"  
Alice added as an explanation.

"Oh, you're sure to do that,"  
said the Cat, "if you only walk long

# Merci de m'avoir écoutée!

Alexane Jouglar