



Annoter, partager et comparer avec Glozz, Inception et TXM-URS

L.-M. Ho-Dac (CLLE, Toulouse), C. Poudat (BCL, Nice)

CORLI, 18 mai 2020

Plan

- 1 Annoter en linguistique
- 2 Des outils pour annoter
- 3 Annoter
- 4 Perspectives et exploitations

Introduction

- 1970 : Corpus Brown Corpus of Standard American English
 - 1 M de mots
 - Premier projet d'annotation morphosyntaxique
- Années 2000 : développements des corpus arborés
 - Puis développement des annotations sémantiques
- Développement de l'informatique et du TAL
 - TAL probabiliste, mobilisation des méthodes d'approvisionnement par la foule (crowdsourcing) comme Amazon mechanical turk

Quelles annotations en linguistique ?

Annotation morphosyntaxique ou grammaticale

- À chaque mot du texte est associée une annotation explicitant la catégorie morphosyntaxique de ce mot : nom commun, pronom, déterminant, etc.
- Pour ce type d'annotation, chaque mot(-forme) du texte est un marquable, une unité

Annotation syntaxique

- Annotation automatique d'un texte avec des étiquettes syntaxiques, par exemple pour les fonctions grammaticales (sujet, objet) voire pour les constituants de chaque phrase : syntagme nominal, syntagme verbal, syntagme prépositionnel, etc.
- Pour ce type d'annotation, les marquables sont des groupes de mots consécutifs
→ Penn Treebank, French Treebank

Quelles annotations en linguistique ?

Annotations discursives

- Coréférence, relations de discours
→ Penn Discourse Treebank, French Discourse Treebank

Annotations sémantiques et pragmatiques

- Depuis les années 2000, beaucoup d'efforts sont faits pour automatiser certaines annotations sémantiques et pragmatiques
 - annotation des émotions, des frames, des entités nommées, des actes de parole, du hate speech
- Annotation manuelle
- Corpus de petite taille
- Évaluation nécessaire avec des scores d'accord inter-annotateurs
- Méthodes d'apprentissage automatique (e.g. deep learning)

Plan

- 1 Annoter en linguistique
- 2 Des outils pour annoter**
- 3 Annoter
- 4 Perspectives et exploitations

Des outils pour annoter

Des outils qui se veulent génériques

- mais qui ont des "restes" issus de leur origine :
 - outil existant d'exploration de corpus
e.g. TXM et Analec → TXM-URS <http://textometrie.ens-lyon.fr/>
 - un projet visant la constitution d'une ressource
e.g. AnnoDis → Glozz <http://www.glozz.org>
 - outil d'annotation enrichi pour de l'active learning
e.g. WebAnno → Inception <https://inception-project.github.io>
- Des principes communs
 - annotation d'objets linguistiques dans des textes digitalisés
 - permettant une AAO ("analyse assistée par ordinateur") pouvant tirer partie de traitements automatiques e.g. étiquetage morpho-syntaxique, parsing, projection de patrons et de lexiques
 - permettant plusieurs couches d'annotation
 - permettant des analyses croisées entre couches d'annotation
 - permettant la constitution de ressources pour la communauté

Des outils pour annoter

Des outils qui se veulent génériques

- mais qui ont des "restes" issus de leur origine :
 - outil existant d'exploration de corpus
e.g. TXM et Analec → TXM-URS <http://textometrie.ens-lyon.fr/>
 - un projet visant la constitution d'une ressource
e.g. AnnoDis → Glozz <http://www.glozz.org>
 - outil d'annotation enrichi pour de l'active learning
e.g. WebAnno → Inception <https://inception-project.github.io>
- Des principes communs
 - **annotation d'objets linguistiques dans des textes digitalisés**
 - permettant une AAO ("analyse assistée par ordinateur") pouvant tirer partie de traitements automatiques e.g. étiquetage morpho-syntaxique, parsing, projection de patrons et de lexiques
 - permettant plusieurs couches d'annotation
 - permettant des analyses croisées entre couches d'annotation
 - permettant la constitution de ressources pour la communauté

Ce qu'annoter implique en linguistique

Procédure typique

- Charger le corpus/texte à annoter
- Délimiter des segments de texte : les unités (**U**)
- Associer des "traits" aux unités selon
- Relier des unités : les relations (**R**)
- Regrouper des relations et des unités dans des structures complexes : les schémas (**S**)

Ce qu'annoter implique en linguistique

Procédure typique

- Charger le corpus/texte à annoter
 - texte brut
 - document word ou pdf
 - texte structurés voire prémarqués e.g. XML (l'annotation donnera lieu à une couche de structuration supplémentaire)
- Délimiter des segments de texte : les unités (**U**)
- Associer des "traits" aux unités selon
- Relier des unités : les relations (**R**)
- Regrouper des relations et des unités dans des structures complexes : les schémas (**S**)

Ce qu'annoter implique en linguistique

Procédure typique

- Charger le corpus/texte à annoter
- Délimiter des segments de texte : les unités (**U**)
 - **Délimitation dans le texte** (annotation intégrée, *in situ* ⇒ le texte source est modifié ce qui empêche l'interopérabilité et le "multicouche".
 - **Délimitation hors le texte** (annotation débarquée, *stand-off*) ⇒ le texte source n'est jamais édité. Chaque annotation est indépendante et n'interfère pas sur le texte source. Permet des analyses croisées.
 - Géolocalisation des bornes de début et de fin au token (TXM - URS), au caractère (Glozz) ou au choix (Inception)
- Associer des "traits" aux unités selon
- Relier des unités : les relations (**R**)
- Regrouper des relations et des unités dans des structures complexes : les schémas (**S**)

Ce qu'annoter implique en linguistique

Procédure typique

- Charger le corpus/texte à annoter
- Délimiter des segments de texte : les unités (**U**)
- Associer des "traits" aux unités selon
- Relier des unités : les relations (**R**)
- Regrouper des relations et des unités dans des structures complexes : les schémas (**S**)

Ce qu'annoter implique en linguistique

Procédure typique

- Charger le corpus/texte à annoter
- Délimiter des segments de texte : les unités (**U**)
- Associer des "traits" aux unités selon
 - un modèle d'annotation prédéfini non modifiable
 - un modèle d'annotation construit dynamiquement
- Relier des unités : les relations (**R**)
- Regrouper des relations et des unités dans des structures complexes : les schémas (**S**)

Des outils pour exploiter/ visualiser des annotations

Prétraitements et posttraitements

- Prémarquer les textes
- Nettoyer les annotations

Croiser des annotations

- Annotations multiples et mesure de l'accord inter-annotateur
- Annotations relevant de différentes couches d'annotation (syntaxe, discours, etc.)

Analyser les annotations

- Statistiques
- Concordances, textométrie

Plan

- 1 Annoter en linguistique
- 2 Des outils pour annoter
- 3 Annoter**
- 4 Perspectives et exploitations

Annoter

Charger les éléments nécessaires à l'annotation

- les Textes bruts ou préparés à annoter
- le modèle d'annotation

Annoter comme surligner

- Délimiter des unités associées à une catégorie, une couleur
- Leur associer des traits

Exemples issus du projet E-Calm, constitution d'un corpus de manuscrits scolaires transcrits, normalisés, étiquetés automatiquement, annotés discursivement

Chargement d'un texte et d'un modèle d'annotation (Glozz)

Le monstre

Dans le nord de la France se trouvait un paysan qui avait une maison blanche. Elle habitait dans cette maison depuis longtemps. Un jour, le paysan vit un lièvre blanc, ^{pendant une nuit} et il voulait en savoir plus, alors, il sortit dehors avec un couteau. Il regarda dans le lièvre et il vit du sang et des os derrière la charnière, il sortit dehors et vit les os. Il s'arrêta dans un lièvre à l'intérieur de la forêt. Il regarda tout le lièvre et vit un grand mort. Il repartit alors. Tout à coup, un grand bruit le fit sursauter, Il se retourna en entendant ce grand bruit. ~~Il se retourna~~ et il vit un monstre ^{à sa} ~~à~~ ^{sa} ~~sa~~ ^{doit} ~~doit~~ ^{griffes} ~~griffes~~ ^{dégoulinées} ~~dégoulinées~~ le sang, il était violet gluant et ^{gouttes} ~~gouttes~~ ^{de} ~~de ^{sa} ~~sa~~ ^{moitié} ~~moitié~~ ^{avait} ~~avait~~ ^{un} ~~un~~ ^{bruit} ~~bruit~~ ^{inimaginable.} ~~inimaginable.~~ Il cria en voyant le monstre mais il le tua.~~

Chargement d'un texte et d'un modèle d'annotation (Glozz)

File Options Import Export Tools Groups Viewers Plugins SandBox ?

Le monstre

Dans le nord de la France se trouvait un paysan dont sa grand-mère qui avait une maison blanche. Elle habitait dans cette maison depuis longtemps. Un jour, le paysan vue un buisson bouger pendant une nuit il ne voulait en savoir plus, alors, il sortit dehors avec un couteau et une torche. Il regarda dans le buisson et il vi du sang et des pas derrière la cloture, il sorti dehors et suivi les pas. Il sarretai dans un buisson à l'interieur de la forêt Il regarda dans le buisson et vue un ours mort. Il reparti alors. Tout à coup un grand bruit le fit sursoter. Il se retourna en entendant ce grand bruit. et il vue un monstre dont ses doit griffe dégoutx deont, il xx vert gluant et a xx : dans mais xx inodaxxt inimaginable. Je cri en voyant le monstre il le tuer. Depuis cette aventure, les enfants ne sortent plus la nuit.

Features		Informations	
Feature name		Feature val	
nature		add	
del		na	
delType		unclear	

Model	As Text	Tools
Annotation model:		
Units		Relations
mod		coreference
lb		
pb		
space		
phraseConsigne		
unclear		
gap		
intervention		
Err_Orthographe		
coref		
refConsigne		
		Schema
		transcri

Normalisation de l'orthographe avec Glozz

File Options Import Export Tools Groups Viewers Plugins Sandbox ?

Le monstre

Dans le nord de la France se trouvait un paysan dont sa grand-mère qui avait une maison blanche. Elle habitait dans cette maison depuis longtemps. Un jour, le paysan vue un buisson bouger pendant une nuit il ne voulait en savoir plus, alors, il sortit dehors avec un couteau et une torche. Il regarda dans le buisson et il vi du sang et des pas derrière la cloture il sorti dehors et suivi les pas. Il sarreta dans un buisson à l'interieur de la forêt Il regarda dans le buisson et vue un ours mort. Il reparti alors. Tout à coup un grand bruit le fit sursoter Il se retourna en entendant ce grand bruit. et il vue un monstre dont ses doit griffe dégoutx deont, il xx vert gluant et modaxx inimaginable. Je cri en voyant le monstre il le tuer. Depuis cette aventure, les enfants ne sortent plus la nuit.

ID=58
 Type=Err Orthographe
 Author=schemin
 LastModifiedBy=schemin

Feature name	Feature value
incertitude	Pas d'incertitude
correction_impossible	Non
correction_1	le tuer
correction_2	il le tuait
correction_3	il l'a tué
correction_4	
correction_5	
correction_6	

Model	As Text	Tools
Annotation model:		
Units	Relations	Schemata
mod lb pb space phraseConsigne unclear gap intervention Err Orthographe coref refConsigne	coreference	transcrit

Charger un autre modèle, ajouter une couche d'annotation

The screenshot shows the Annoter application window. The main text area contains a story about a man in the north of France who finds a monster in a bush. Several phrases in the text are enclosed in rectangular boxes, representing annotations. The right-hand sidebar is divided into two main sections: 'Features' and 'Model'. The 'Features' section shows a table with feature names and their values. The 'Model' section shows the current annotation model and a list of relations.

File Options Import Export Tools Groups Viewers Plugins Sandbox ?

Le monstre

Dans le Nord de la France se trouvait un paysan dont sa grand-mère qui avait une maison blanche. Elle habitait dans cette maison depuis longtemps. Un jour, le paysan vit un buisson bouger pendant une nuit, il voulait en savoir plus, alors, il sortit dehors avec un couteau et une torche. Il regarda dans le buisson et il vit du sang et des pas derrière la clôture, il sortit dehors et suivit les pas. Il s'arrêta dans un buisson à l'intérieur de la forêt. Il regarda dans le buisson et vit un ours mort. Il repartit alors. Tout à coup un grand bruit le fit sursauter. Il se retourna en entendant ce grand bruit. Et il vit un monstre dont ses doigts griffés dégoulaient de sang, il était vert gluant et n'avait pas d'œils mais avait un odorat inimaginable. Je crie en voyant le monstre mais il le tua. Depuis cette aventure, les enfants ne sortent plus la nuit.

Features		Informations
Feature name	Feature value	
type	maillon_Elle	
groupe	Non	
incertitude sur la délimitation	Non	
incertitude sur le rattachement	Non	
commentaire		

Model	As Text	Tools
Annotation model:		
Units		
maillon-candidat		
maillon_Elle		
maillon_Il		
maillon_lesEnfants		
Relations		
coref_Elle		
coref_Il		
coref_lesEnfants		
Schemas		

Charger un autre modèle, ajouter une couche d'annotation

File Options Import Export Tools Groups Viewers Plugins Sandbox ?

The screenshot shows the Annoter application interface. The main window displays a text document titled "Le monstre" with several annotations. A toolbar at the top contains various icons for file operations and editing. On the right side, there are two panels: "Features" and "Model".

Text Document:

Le monstre

Dans le Nord de la France se trouvait un paysan dont sa grand-mère qui avait une maison blanche. Elle habitait dans cette maison depuis longtemps. Un jour, le paysan vit un buisson bouger pendant une nuit, il voulait en savoir plus, alors, il sortit dehors avec un couteau et une torche. Il regarda dans le buisson et il vit du sang et des pas derrière la clôture, il sortit dehors et suivit les pas. Il s'arrêta dans un buisson à l'intérieur de la forêt. Il regarda dans le buisson et vit un ours mort. Il repartit alors. Tout à coup un grand bruit le fit sursauter, il se retourna en entendant ce grand bruit. Et il vit un monstre dont ses doigts griffés dégoulaient de sang, il était vert gluant et n'avait pas d'yeux mais avait un odorat inimaginable. Je crie en voyant le monstre mais il le tua. Depuis cette aventure, les enfants ne sortent plus la nuit.

Features Panel:

Feature name	Feature value
type	maillon_Elle
groupe	Non
incertitude sur la délimitation	Non
incertitude sur le rattachement	Non
commentaire	

Model Panel:

Annotation model:	Relations	Schemas
maillon-candidat	coref_Elle	
maillon_Elle	coref_Il	
maillon_Il	coref_lesEnfants	
maillon_lesEnfants		

Besoins à l'origine de Glozz : projet ANNODIS, annotation discursive

Spécificité des textes et objets à annoter

- Textes longs et structurés (structures de haut niveau, prise en compte de la structure de document)
- Structures complexes, qui peuvent se "chevaucher" et sont influencées par la structure du document

Besoins

- Avoir une vue de haut du texte => ruban
- Avoir une vue "document" des textes à annoter
- Permettre de cibler vers des zones à annoter => prémarquage
- Avoir une vue graphique des annotations

Besoins à l'origine de Glozz : projet ANNODIS, annotation discursive

Spécificité des textes et objets à annoter

- Textes longs et structurés (structures de haut niveau, prise en compte de la structure de document)
- Structures complexes, qui peuvent se "chevaucher" et sont influencées par la structure du document

Besoins

- **Avoir une vue de haut du texte => ruban**
- **Avoir une vue "document" des textes à annoter**
- Permettre de cibler vers des zones à annoter => prémarquage
- Avoir une vue graphique des annotations

Vue de haut – ruban – et vue réaliste du "document"

(actuellement colonel) vient d'être remplacé par Sam Ringer. Ce dernier, déjà en poste au Pentagone sous Clinton, est plus senior. Il est connu pour sa force de caractère et tiendra sans doute à reprendre l'autorité sur les sous-groupes du PCC espace.

Pour les autres cellules de la Maison Blanche, on note les évolutions suivantes :

** Le rôle des conseillers économiques devrait être moindre dans la formulation de la politique spatiale que dans les années précédentes. Sous l'Administration Clinton les questions spatiales ont souvent été traitées sous un angle économique (imposition de standards techniques, choix des politiques d'exportation, décisions sur la navigation et sur l'observation commerciale) et les responsables du National Economic Council avaient une forte influence dans les décisions. Désormais, l'espace est vu sous un angle plus clairement militaire.

1) L'Office of Scientific and Technology Policy (OSTP) conserve son rôle d'encadrement des décisions relatives à la Nasa. Le poste d'assistant du Directeur pour les questions " Espace et Aéronautique " est actuellement vacant. Scott Pace, nommé par Bush, vient de quitter ces fonctions pour rejoindre le bureau de l'administrateur de la Nasa.

Comme pour le NSC, la pratique de la nouvelle administration a été jusqu'à présent de ne pas faire présider par l'OSTP les groupes de réflexion auxquels il participe. Il est possible que le départ de Scott Pace ait à voir avec ce mode de fonctionnement peu efficace.

B - Les programmes spatiaux militaires pour 2003

Dans le cadre de la réorganisation de ses services, Donald Rumsfeld a également lancé des réformes favorables à une meilleure gestion des

Besoins à l'origine de Glozz : projet ANNODIS, annotation discursive

Spécificité des textes et objets à annoter

- Textes longs et structurés (structures de haut niveau, prise en compte de la structure de document)
- Structures complexes, qui peuvent se "chevaucher" et sont influencées par la structure du document

Besoins

- Avoir une vue de haut du texte => ruban
- Avoir une vue "document" des textes à annoter
- Permettre de cibler vers des zones à annoter => prémarquage
- Avoir une vue graphique des annotations

Besoins à l'origine de Glozz : projet ANNODIS, annotation discursive

Spécificité des textes et objets à annoter

- Textes longs et structurés (structures de haut niveau, prise en compte de la structure de document)
- Structures complexes, qui peuvent se "chevaucher" et sont influencées par la structure du document

Besoins

- Avoir une vue de haut du texte => ruban
- Avoir une vue "document" des textes à annoter
- **Permettre de cibler vers des zones à annoter => prémarquage**
- Avoir une vue graphique des annotations

Vue de haut de zones à annoter (concentration de prémarques ou zone non annotée)

meuvent colonel vont être remplacé par Carl Kingler. Ce dernier, déjà en poste au Pentagone sous Clinton, est plus senior. **III** est connu pour sa force de caractère et tiendra sans doute à reprendre l'autorité sur les sous- groupes du PCC espace.

Pour les autres cellules de la Maison Blanche, on note les évolutions **suivantes :**

** **Le** rôle des conseillers économiques devrait être moindre dans la formulation de la politique spatiale que dans les années précédentes. Sous l'Administration Clinton **les questions** spatiales ont souvent été traitées sous un angle économique (imposition de standards techniques, choix des politiques d'exportation, décisions sur la navigation et sur l'observation commerciale) et les responsables du National Economic Council avaient une forte influence dans les décisions. **Désormais** **l'espace** est vu sous un angle plus clairement militaire.

1) **L'**Office of Scientific and Technology Policy (OSTP) conserve son rôle d'encadrement des décisions relatives à la Nasa. **Le poste** d'assistant du Directeur pour les questions " Espace et Aéronautique " est actuellement vacant. Scott Pace, nommé par Bush, vient de quitter ces fonctions pour rejoindre le bureau de l'administrateur de la Nasa.

Comme pour le NSC, la pratique de la nouvelle administration a été jusqu'à présent de ne pas faire présider par l'OSTP les groupes de réflexion auxquels il participe. Il est possible que le départ de Scott Pace ait à voir avec ce mode de fonctionnement peu efficace.

B - Les programmes spatiaux militaires pour 2003

Dans le cadre de la réorganisation de ses services, **Donald Rumsfeld** a également lancé des réformes favorables à une meilleure gestion des

Annotation localement et largement, indépendamment de la structure de document

Annoter des structures complexes et permettre des chevauchements <p> et <structure>

The screenshot displays a document with several paragraphs of text. The text is annotated with various colored boxes and callouts:

- A green box highlights the sentence: "Le nouveau colonel venant de remplacer par son prédécesseur le poste au Pentagone sous Clinton, est plus senior, il est connu pour sa force de caractère et tiendra sans doute à reprendre l'autorité sur les sous-groupes du PCC espace." A red dot on the left is connected to this box by a dotted line.
- A red box highlights the sentence: "Pour les autres cellules de la Maison Blanche, on note les évolutions suivantes:"
- A yellow box highlights a paragraph: "Le rôle des conseillers économiques devrait être moindre dans la formulation de la politique spatiale que dans les années précédentes. Sous l'Administration Clinton les questions spatiales ont souvent été traitées sous un angle économique (imposition de standards techniques, choix des politiques d'exportation, décisions sur la navigation et sur l'observation commerciale) et les responsables du National Economic Council avaient une forte influence dans les décisions. Désormais, l'espace est vu sous un angle plus clairement militaire." A blue dot on the right is connected to this box by a dotted line.
- A yellow box highlights a numbered list item: "1) L'Office of Scientific and Technology Policy (OSTP) conserve son rôle d'encadrement des décisions relatives à la Nasa. Le poste d'assistant du Directeur pour les questions "Espace et Aéronautique" est actuellement vacant. Scott Pace, nommé par Bush, vient de quitter ces fonctions pour rejoindre le bureau de l'administrateur de la Nasa." A blue dot on the right is connected to this box by a dotted line.

Below the annotated text, there is a paragraph of text:

Comme pour le NSC, la pratique de la nouvelle administration a été jusqu'à présent de ne pas faire présider par l'OSTP les groupes de réflexion auxquels il participe. Il est possible que le départ de Scott Pace ait à voir avec ce mode de fonctionnement peu efficace.

Inception

Active Learning

Session

Layer: Named entity

Terminate

Recommendation

Text: Illinois

Label: LOC

Score: 1

Delta: 1

Accept Reject Skip

Learning History

...	skipped	
Berkeley	http://www.wikidata.org/entity/Q1168756	skipped
Tesla	PER	accepted
Tesla	PER	accepted
Tesla	PER	accepted
Tesla	PER	accepted
Tesla	PER	accepted
Science	OTH	rejected
Tesla	PER	accepted

Annotation

1 Barack Hussein Obama II born August 4, 1961) is an American politician who served as the 44th President of the United States from 2009 to 2017 .

2 The first African American to assume the presidency, he was previously the junior United States Senator from Illinois from 2005 to 2008.

3 He served in the Illinois State Senate from 1997 until 2004.

Illinois Senate
upper chamber of the Illinois General Assembly, the legislative branch of the government of the state of Illinois in the United States

Layer: Surface form

Annotation: Delete Clear

Layer: Named entity

Text: Illinois

Identifier: Illi

- Illinois
- Illinois Senate
- Illinois River
- Governor of Illinois
- Alton
- Illinois Country
- Illinois Territory

TXM – URS

Desperiers - 1

Unité: Maillon : Maillon-3 : Créer Supprimer

et le **povre** **Caillette** demouroit
là, et ne **disoit** mot: Car il n'avoit point
d'autre apprehension, sinon qu'il pensoit
estre confiné là pour toute sa vie. Il passe **un des**
Seigneurs de court, qui le voit ainsi en conseil avec ce
pillier, qui le fait incontinent desgager de là: s'enquerant
bien expressément qui avoit fait celà, et qui l'ha
mis là? Que voulez **vous**, **un sot** l'ha mis là, **un sot** l'ha
mis. Ouand **on** disoit. Ce ont esté **les pages**. **Caillette**

default [Navigation] 1 / 5 [Navigation]

ChainesMacro.groovy DESPERIERS:([id="w_Desperiers_17"] [id="w_Desperiers_18"])([id="w_Desperiers_27"] [id="w_Desperiers_28"])

[Navigation] 1 - 32 / 32 [Navigation]

text_id	Contexte gauche	Pivot	Contexte droit
Desperiers	et Polite. LES pages avoyent attaché l'oreille	à Caillette	avec un clou contre un posteau,
Desperiers	avec un clou contre un posteau, et	le povre Caillette	demeuroit là, et ne disoit mot: c
Desperiers	le povre Caillette demouroit là, et ne	disoit	mot: Car il n'avoit point d'autre :
Desperiers	là, et ne disoit mot: Car	il	n'avoit point d'autre apprehensi
Desperiers	Car il n'avoit point d'autre apprehension, sinon	qu'il	pensoit estre confiné là pour toi
Desperiers	sinon qu'il pensoit estre confiné là pour toute	sa	vie. Il passe un des Seigneurs de
Desperiers	passé un des Seigneurs de court. qui	le	voit ainsi en conseil avec ce pillie

26 to 28 Référentiels

Catégorie: GN Défini Expansion NP Fonction Sujet

Comparaison des 3 outils

	Glozz	Inception	TXM
Gestion de (sous-)corpus	1 texte -- 1 fichier 	corpus et textes 	sous-corpus et partitions 
Visualisation ergonomique	annoter un document 	annoter un fichier (une "phrase" par ligne) 	annoter un texte 
Prémarquage	en prétraitement 	Intégré (active learning - différents niveaux) 	contraint et en prétraitement 
Annotation d'unités	délimitation de l'unité au caractère 	délimitation de l'unité au choix 	délimitation de l'unité au token 
de relations entre unités			
de structures complexes			

Plan

- 1 Annoter en linguistique
- 2 Des outils pour annoter
- 3 Annoter
- 4 Perspectives et exploitations

Perspectives et exploitations

- Quelles fonctionnalités retenir ?
- Outils pour la multi-annotation
 - Outils pour la gestion de plusieurs annotateurs (Inception)
 - Calcul de l'accord inter-annotateur intégré
 - Outils d'adjudication (Glozz)
- Modules d'analyse des données annotées (Glozz en SQL, TXM avec des macros, Inception et les modules d'active learning intégrés)
- Méthodes de la linguistique de corpus (TXM, les origines)
- Formats d'import-export (JSON, XML) et l'importance de l'unité de géolocalisation
 - Nécessité de "mettre les mains dedans"
 - Des décisions non triviales et au delà des aspects techniques