

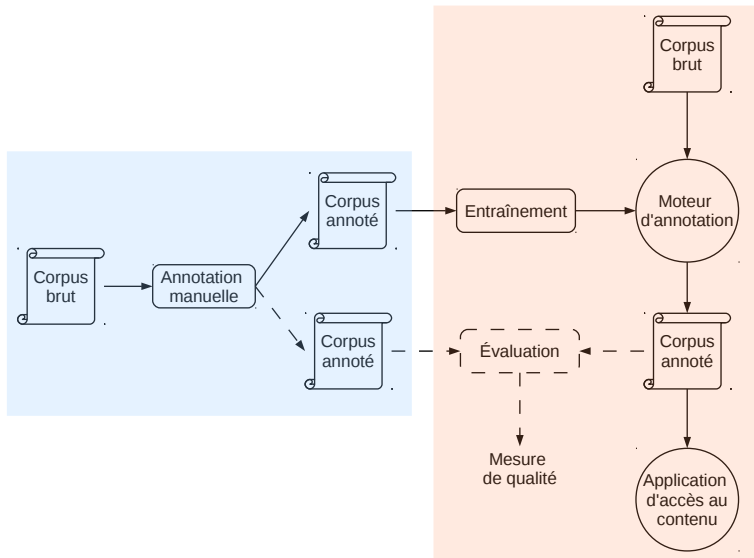
Les ressources annotées, un enjeu pour l'analyse de contenu : vers une méthodologie de l'annotation manuelle de corpus

Karën Fort

7 décembre 2012



Des corpus annotés en TAL, pourquoi faire ?



Le renouveau des corpus

Date	Nb études sur corpus
Jusqu'à 1965	10
1966-1970	20
1971-1975	30
1976-1980	80
1981-1985	160
1985-1991	320

[Johansson, 1991]

Après une stagnation :

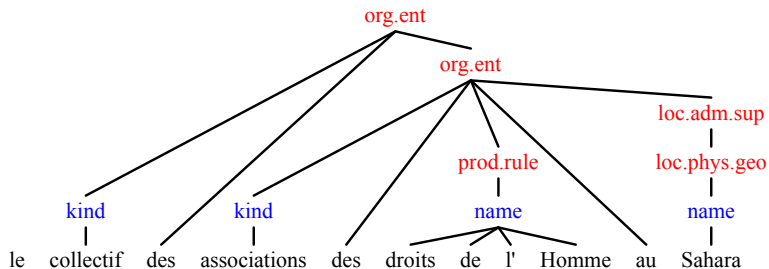
- de plus en plus de corpus
- de plus en plus de corpus **annotés** :
 - ▶ de plus en plus **gros**
 - ▶ de plus en plus **divers**
 - ▶ de plus en plus **complexes**

Complexification et diversification

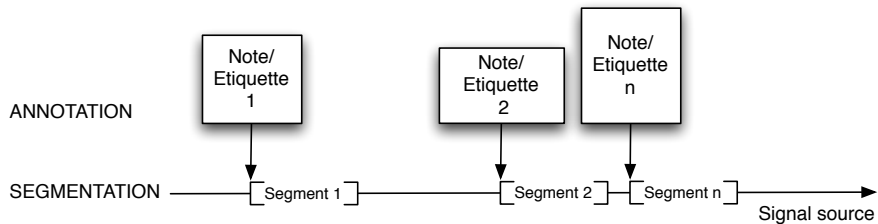
Morpho-syntaxe [Marcus et al., 1993] :

I/PRP do/VBP n't/RB feel/VB very/RB ferocious/JJ ./.

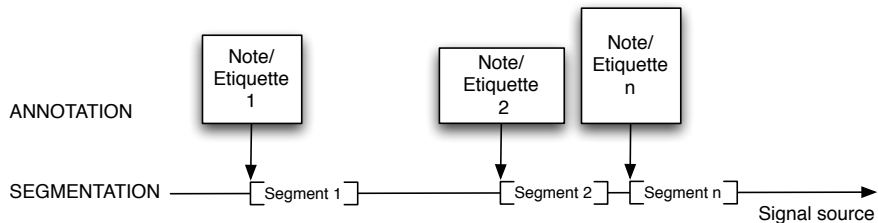
Entités nommées structurées [LAW 2011] :



Annotation



Annotation



Ajout d'informations **interprétatives** [Leech, 1997, Habert, 2005]

Annotation

Une vision large :

- **nature** du support : texte, vidéo, images, etc.
- **contenu** sémantique : note chiffrée, valeur choisie dans un référentiel fermé ou texte libre
- **positionnement** : global ou local
- **objectif** : visée évaluative ou caractérisante, simple commentaire

Des expériences :

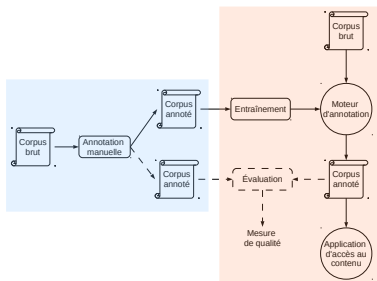
- sur du **texte** (source ou transcriptions),
- avec un référentiel fermé
- avec un positionnement local

L'application : horizon de l'annotation

Une annotation est toujours orientée par une tâche [Habert, 2000].

- ★ visée applicative directe (résumés de matchs pour la campagne football)
- ★ application intermédiaire ou interne au TAL (étiquetage morpho-syntaxique)

Une annotation est d'autant plus utile qu'elle a été conçue en fonction d'une application spécifique [Leech, 2005].



Exemple : entités nommées en microbiologie [TALN 2009]

We conclude that 3CDproM can process both structural and nonstructural precursors of the poliovirus polyprotein and that it is active against a synthetic peptide substrate.

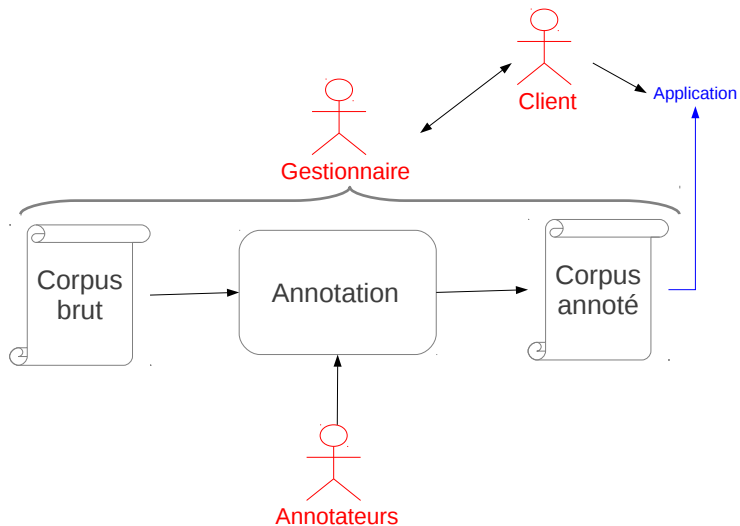
→ Annotation d'indexation :

We conclude that `<gene>3CDproM</gene>` can process both structural and non-structural precursors of the `<EukVirus uncertainty-type="too-generic">poliovirus polyprotein</EukVirus>` and that it is active against a synthetic peptide substrate.

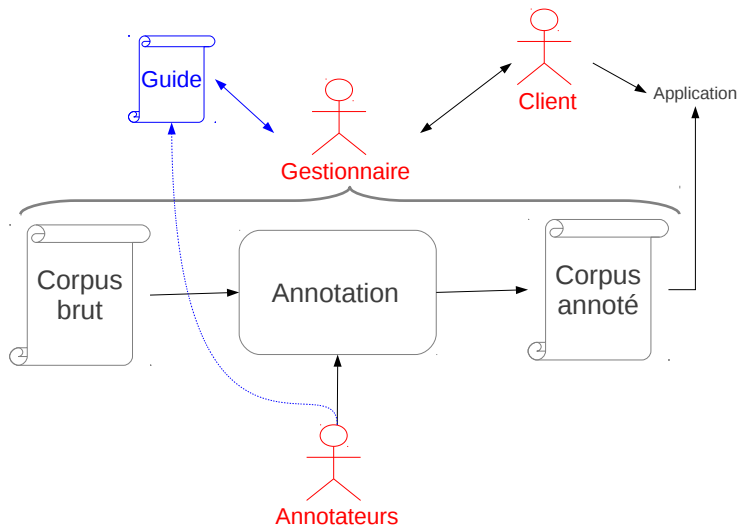
→ Annotation de modélisation :

We conclude that `<EukVirus>3CDproM</EukVirus>` can process both structural and nonstructural precursors of the `<EukVirus uncertainty-type="too-generic"><taxon>poliovirus</taxon>polyprotein</EukVirus>` and that `<EukVirus>it</EukVirus>` is active against a synthetic peptide substrate.

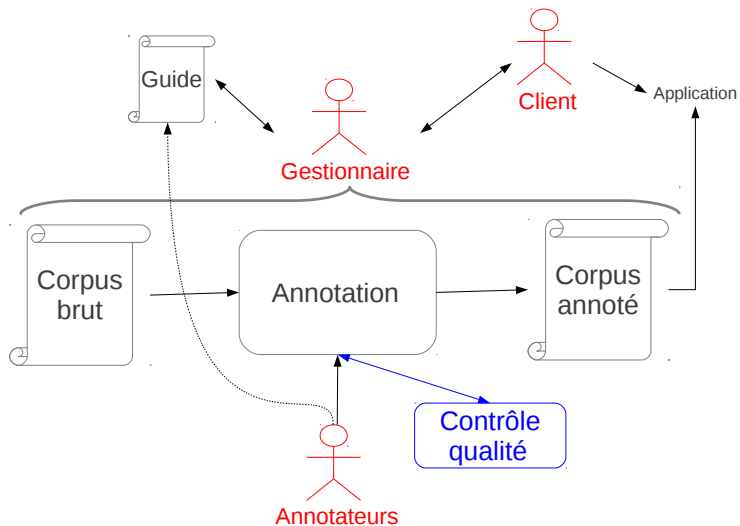
Campagne d'annotation



Campagne d'annotation



Campagne d'annotation



- 1 Introduction
- 2 Une maîtrise insuffisante du processus d'annotation
- 3 Organiser une campagne d'annotation
- 4 Analyser la complexité d'une campagne d'annotation
- 5 Évaluer les annotations manuelles résultantes
- 6 Conclusion et perspectives

- 1 Introduction
- 2 Une maîtrise insuffisante du processus d'annotation
 - Le coût : question centrale
 - Quelques solutions proposées
 - Une méthodologie à construire
- 3 Organiser une campagne d'annotation
- 4 Analyser la complexité d'une campagne d'annotation
- 5 Évaluer les annotations manuelles résultantes
- 6 Conclusion et perspectives

Un coût élevé, voire prohibitif

Penn Treebank [Marcus et al., 1993] :

- 4,8 millions de tokens annotés en morpho-syntaxe \Rightarrow phase d'apprentissage d'1 mois, pour atteindre 3 000 mots/h
- 3 millions de tokens annotés en syntaxe \Rightarrow phase d'apprentissage d'2 mois, pour atteindre 475 mots/h

Prague Dependency Treebank [Böhmová et al., 2001] :

- 1,8 millions de tokens annotés en morpho-syntaxe et syntaxe
 \Rightarrow 5 ans, 22 personnes (max. 17 en parallèle), 600 000 dollars

GENIA [Kim et al., 2008] :

- 9 372 phrases annotés en microbiologie (noms de gènes, protéines)
 \Rightarrow 5 annotateurs à temps partiel, 1 coordinateur senior et 1 junior pendant 1,5 an

Des solutions parcellaires

- Outiller l'annotation :
 - ▶ outils d'aide à l'annotation
 - ▶ annotation partiellement automatisée
- Faire annoter des “non-experts”
- Appliquer une méthodologie partielle

Outils d'aide à l'annotation



Outils d'aide à l'annotation

The screenshot shows a software interface for text annotation. The main window displays a document with various text elements highlighted in different colors (yellow, green, red, blue). A table overlay is visible in the center, listing different annotation styles. The table has three columns: 'Unit style', 'Relation style', and 'Schema style'. The 'Unit style' column lists various actions and entities like 'ActionDuPublic', 'ArreterBut', 'Centrer', etc. The 'Relation style' column lists 'Background-color'. The 'Schema style' column has a 'Hide' checkbox. To the right, there is a sidebar with a search box, a list of 'Units', 'Relations', and 'Schemas', and a table with columns 'Sort/Type', 'Sort/Date', 'Show sel.', and 'Visible'.

Unit style	Relation style	Schema style
Type name	Background-color	Hide
A.ActionDuPublic		
A.ArreterBut		
A.Centrer		
A.Dribbler		
A.FaireFautDejeu		
A.FaireTentative2c...		
A.Horsjeu		
A.IntercepterBallon		
A.MarqueBut		
A.PasserBallon		
A.PrendreCartonJaune		
A.PrendreCartonRou...		
A.PrendreRappelLo...		
A.PasserBut		
A.TirerCorner		
A.TirerCoupFrancDir...		
A.TirerCoupFrancInd...		
A.TirerPenalty		
ActionPourActeurVide		
Arbitre		
ArbitreAssistant		
Entraîneur		
Equipe		
EspaceSurTerrain		
loueur		

90+3
Les **Bordeaux** ont effectué un super match et son rivaliser face à un **Chelsea** solide mais peu réaliste avant de se déplacer à Rome.

90+3
[Unité annotée lors de la vérification] un espace **loueur** en faveur [Unité ajoutée lors de la phase ann] Le ballon est **reposé** en extrémités de la surface.

90
[Temps additionnel] 3 minutes.

89
Belle tentative de **Carriag** qui manque le **cadre**.

87
Lampard est expulsé après avoir reçu son deuxième carton jaune.

87
L'International anglais, **Franck Lampard** ~~écoupe d'un carton jaune pour avoir effectué un tir~~ [Unité ajoutée lors de la phase annotation des relations] appuie.

85
[Unité supprimée lors de la phase annotation des actions] Paulo était en doublon avec Paulo Ferreira.
Paulo Ferreira entre en jeu à la place de son coéquipier anglais **Joe Cole**.

84
Chamak écoupe d'un carton jaune.

83
[Un] **La comar** désigne l'action, mais aussi un [Vérification dans la fiche du match que Girard a bien marqué un but très beau] **corrier** de **Couruff** qui ~~écoupe le ballon d'arbitrage sur le côté du capitaine bordeaux~~ **Yllo** qui n'a plus qu'à ajuster la trajectoire de la balle et redonne un espoir de qualification aux **Girondins**.

81
Les **Girondins** conservent le ballon et se rapprochent timidement vers la **surface asglaire**.

78
Belle tentative de l'International anglais, **Franck Lampard**, directement sur **Valverde** qui repousse des poings.

[Widlöcher and Mathet, 2009]

L'impossible outil à tout faire

Pas d'outil faisant l'unanimité :

- **diversité** des annotations \Rightarrow diversité des outils d'aide à l'annotation [Cunningham et al., 2002, de la Clergerie, 2008, Felt et al., 2010, Kaplan et al., 2010, Landragin et al., 2012]
- coût de la **phase d'apprentissage** \Rightarrow utilisation d'outils "maison", souvent très simples mais limités (plugins XEmacs)

Annotation assistée par ordinateur

Propagation d'étiquettes :

- ré-utiliser automatiquement les annotations déjà réalisées :
 - + très simple à mettre en œuvre
 - efficacité dépendant du nombre d'annotations déjà réalisées
 - biais [Carmen et al., 2010] : les étiquettes proposées sont privilégiées

Annotation assistée par ordinateur

Pré-annotation automatique :

- annotation réalisée de manière automatique par un outil (de TAL), puis **corrigée** par l'humain :
 - + gain en temps et en cohérence [Marcus et al., 1993] **quand** la pré-annotation est de qualité
 - biais [TALN 2009] : les éléments non pré-annotés sont négligés
 - quelle pré-annotation pour quels gains en **temps** et en **qualité** d'annotation ?
 - évaluation exacte du **biais** ?

Collaboration en ligne : faire annoter les “non-experts”

Jeux “ayant un but” (GWAP) [von Ahn, 2006] :

- tâche “déguisée” derrière une interface ludique :
 - + beaucoup d'annotations en peu de temps (100 000 relations anaphoriques en 2008 [Chamberlain et al., 2008])
 - **coût** élevé de conception du jeu et de l'interface
 - gestion de la **tricherie**
 - qualité difficile à évaluer

Myriadisation du travail parcellisé (*microworking* à la Amazon Mechanical Turk) :

- tâches découpées en sous-tâches et réalisées par des “non-experts”
 - met directement l'accent sur la **rémunération**
 - pas une solution réelle (difficulté d'évaluer la qualité et problèmes éthiques) [Comp. Ling. 2011]

Méthodologies partielles

[Bonneau-Maynard et al., 2005] :

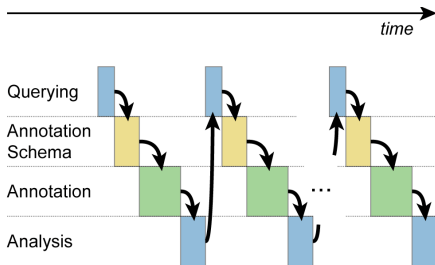
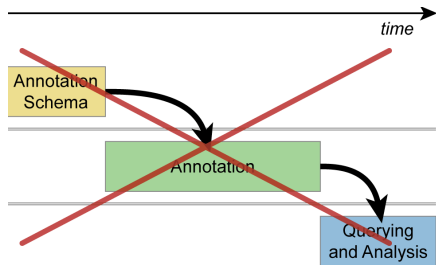
- calculer l'accord inter-annotateurs très tôt en début de campagne
- modifier et améliorer le guide d'annotation en fonction de ces retours

Méthodologies partielles

[Bonneau-Maynard et al., 2005] :

- calculer l'accord inter-annotateurs très tôt en début de campagne
- modifier et améliorer le guide d'annotation en fonction de ces retours

Annotation agile [Voormann and Gut, 2008] : plusieurs aller-retours entre l'annotation et la mise à jour du guide :



Une méthodologie globale qui reste à construire

Un existant **parcellaire** et mal documenté :

- bonnes pratiques de **haut niveau**
- solutions au **cas par cas** : biais ? qualité ? éthique ?
- **brèves** de méthodologies
- mesures d'évaluation **pas toujours adaptées** et peu explicitées

Une méthodologie globale qui reste à construire

Un existant **parcellaire** et mal documenté :

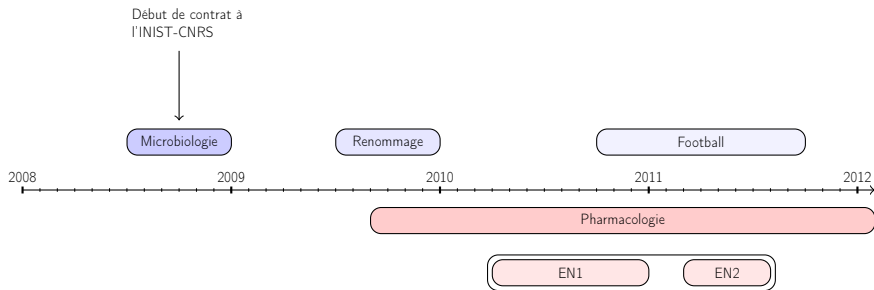
- bonnes pratiques de **haut niveau**
- solutions au **cas par cas** : biais ? qualité ? éthique ?
- **bribes** de méthodologies
- mesures d'évaluation **pas toujours adaptées** et peu explicitées

Nécessité d'une vision **d'ensemble** :

- prise en compte des **acteurs**, en particulier des annotateurs
- **formalisation** de l'annotation
- mesures d'évaluation **adaptées**

- 1 Introduction
- 2 Une maîtrise insuffisante du processus d'annotation
- 3 Organiser une campagne d'annotation**
 - Participation à des campagnes d'annotation
 - Organiser une campagne d'annotation
 - Processus en œuvre lors d'une pré-campagne
- 4 Analyser la complexité d'une campagne d'annotation
- 5 Évaluer les annotations manuelles résultantes
- 6 Conclusion et perspectives

Participation aux campagnes d'annotation Quæro



Noms de gènes, de protéines et d'espèces [TALN 2009]

- client : INRA MIG
- corpus : 110 000 tokens issus de MEDLINE (résumés en anglais)
- objectif : entraîner l'outil RenBio

keratinocyte growth factor receptor (KGFR)

- **correction** de pré-annotations automatiques
 - 2 annotateurs experts en microbiologie (INIST)
 - travail en **tandem**
 - outil d'annotation Cadix [Alphonse et al., 2004]
- **Pas de réelle gestion** de l'annotation
- Annoter en fonction d'un outil (RenBio) ?

Relations de renommage de gènes [BioNLP 2011]

- client : INRA MIG
- corpus : 400 000 tokens issus de MEDLINE (résumés en anglais)
- objectif : entraîner et évaluer les outils de la "shared task" BioNLP

The yppB gene complemented the defect of the recG40 strain. yppB and ypbC and their respective null alleles were termed recU and "recU1" (recU:cat) and recS and "recS1" (recS:cat), respectively.

- 2 annotateurs experts en microbiologie (INIST)
 - gestionnaire de l'annotation
 - outil d'annotation Cadix
- Application de la méthodologie de [Bonneau-Maynard et al., 2005]
- Qualité difficile à évaluer, mais satisfaisante ($\kappa = 0,8$) [LREC 2012a]

Renommage

2010

Entités nommées, actions et relations en football

[Corpus Linguistics 2011]

- client : INRIA TexMex
- corpus en français, pré-annoté en entités nommées, hétérogène
- application : résumé automatique de matchs de football

C'est au tour de Quintin de recevoir un jaune pour un tacle très appuyé sur Traoré.

- gestionnaire de l'annotation
 - 2 annotateurs experts en football (INIST)
 - outil d'annotation Glozz
 - 3 couches d'annotation (unités, actions, relations)
- **Pré-campagne**, création d'une **mini-référence**, évaluation
- Annotation complexe, résultats hétérogènes, corrections nécessaires
- [LREC 2012]

Entités nommées, termes et relations en pharmacologie

- client : Jouve
- corpus : brevets (en anglais) de tailles indéterminées (110 pour l'instant, soit 20 Mo)
- application : aide au processus d'examen des brevets

Topoisomerase-inhibitors are useful for the treatment of inflammatory disorders including arthritis, restonosis, surgical adhesions and other diseases. Controlled-release polymeric formulations to topoisomerase-inhibitors are particularly suitable for this use.¶

- 2 annotatrices expertes en pharmacologie (INIST)
 - gestionnaire de l'annotation au début, puis transfert
 - outil d'édition de XML Epic
 - au début entités nommées et termes, puis ajout de relations sémantiques de type "dérivé_de"
- **Pré-campagne**, création d'une **mini-référence**, évaluation
- Application mal définie



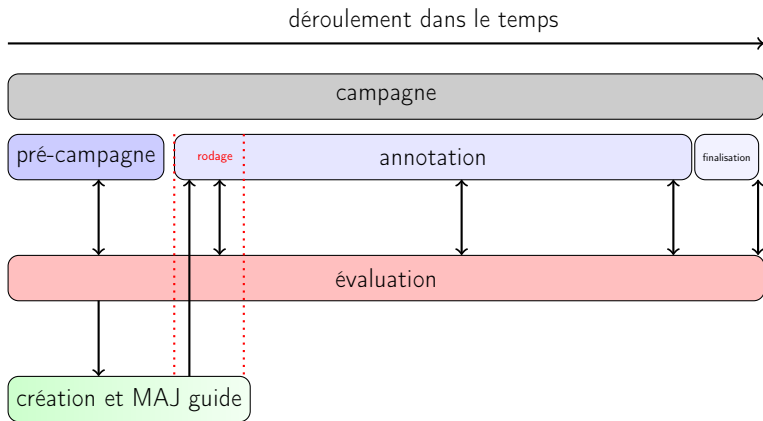
Entités nommées structurées [LAW 2012]

- presse orale (EN1) et ancienne (EN2) en français
- client : membres de Quæro (campagne d'évaluation)
- corpus : 1,2 millions de tokens transcrits (EN1) et 1,7 millions de tokens (EN2)
- objectif : entraînement et évaluation des outils Quæro

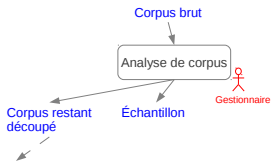
*football toujours, le <pers.coll><loc.adm.nat> Brésil
</loc.adm.nat></pers.coll> n'a pas le droit à l'erreur ce soir au
<loc.fac><kind> stade </kind><name>Gerland </name></loc.fac> .*

- 4 annotateurs encadrés par un responsable (ELDA)
 - gestionnaire de l'annotation : LIMSI
 - outil XEmacs avec plugin
- Création d'une **mini-référence**, annotation en double (EN1), évaluation
- Utilisation pour la campagne ETAPE

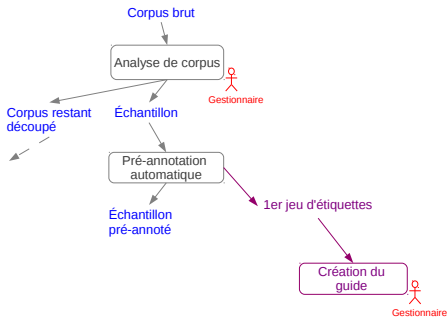
Vue d'ensemble d'une campagne d'annotation



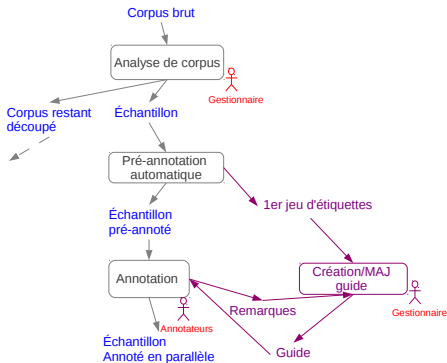
Processus en œuvre lors de la pré-campagne



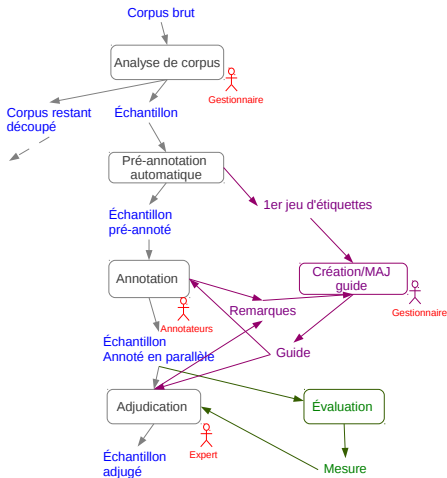
Processus en œuvre lors de la pré-campagne



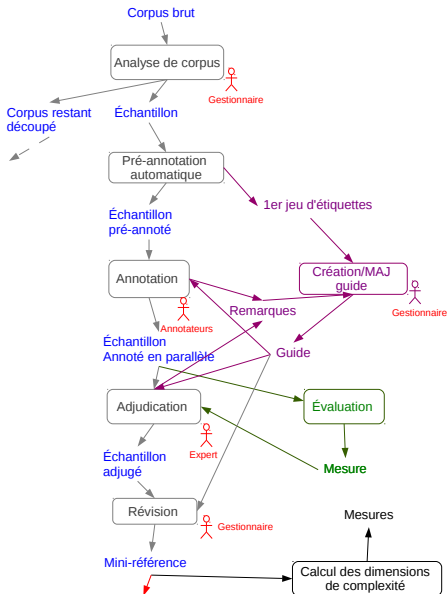
Processus en œuvre lors de la pré-campagne



Processus en œuvre lors de la pré-campagne



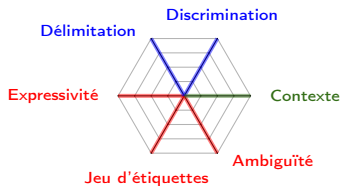
Processus en œuvre lors de la pré-campagne



- 1 Introduction
- 2 Une maîtrise insuffisante du processus d'annotation
- 3 Organiser une campagne d'annotation
- 4 Analyser la complexité d'une campagne d'annotation**
 - Quoi annoter ?
 - Comment annoter ?
 - Le poids du contexte
 - Synthèse
- 5 Évaluer les annotations manuelles résultantes
- 6 Conclusion et perspectives

Dimensions de complexité [COLING 2012]

- 5 dimensions indépendantes :
 - ▶ 2 liées à la **localisation** des annotations
 - ▶ 3 liées à la **caractérisation** des annotations
- 1 non indépendante : le **contexte**



- Échelle de **0** (complexité nulle) à **1** (complexité maximale) pour permettre la comparaison
- Indépendantes du volume à annoter et du nombre d'annotateurs impliqués

Tâche d'Annotation Élémentaire (TAE)

Passer d'une tâche complexe à plusieurs tâches simples :

Critère

Une tâche d'annotation peut être décomposée en au moins deux TAE si son jeu d'étiquettes est décomposable en jeux d'étiquettes plus réduits et indépendants

→ peut correspondre à plusieurs phases d'annotation successives ou pas

Illustration : renommage de gènes

- 1 Identification des noms de gènes dans le signal source :

*The **yppB** gene complemented the defect of the recG40 strain. **yppB** and **y pbC** and their respective null alleles were termed “**recU**” and “**recU1**” (recU:cat) and “**recS**” and “**recS1**” (recS:cat), respectively.*

- 2 Identification des couples de gènes exprimant une relation de renommage :

*The **yppB** gene complemented the defect of the recG40 strain. **yppB** and **y pbC** and their respective null alleles were termed “**recU**” and “**recU1**” (recU:cat) and “**recS**” and “**recS1**” (recS:cat), respectively.*

Discrimination

Parties du discours [Marcus et al., 1993], pré-annotées :

I/PRP do/VBP n't/RB feel/VB very/RB ferocious/JJ ./.

Renommage de gènes [LREC 2012a], non pré-annoté :

The yppB:cat and ypbC:cat null alleles rendered cells sensitive to DNA-damaging agents, impaired plasmid transformation (25- and 100-fold), and moderately affected chromosomal transformation when present in an otherwise Rec+ B. subtilis strain. The yppB gene complemented the defect of the recG40 strain. yppB and ypbC and their respective null alleles were termed recU and "recU1" (recU:cat) and recS and "recS1" (recS:cat), respectively. The recU and recS mutations were introduced into rec-deficient strains representative of the alpha (recF), beta (addA5 addB72), gamma (rech342), and epsilon (recG40) epistatic groups.

Discrimination

Parties du discours [Marcus et al., 1993], pré-annotées :

I/PRP do/VBP n't/RB feel/VB very/RB ferocious/JJ ./.

Renommage de gènes [LREC 2012a], non pré-annoté :

The yppB:cat and ypbC:cat null alleles rendered cells sensitive to DNA-damaging agents, impaired plasmid transformation (25- and 100-fold), and moderately affected chromosomal transformation when present in an otherwise Rec+ B. subtilis strain. The yppB gene complemented the defect of the recG40 strain. yppB and ypbC and their respective null alleles were termed recU and "recU1" (recU:cat) and recS and "recS1" (recS:cat), respectively. The recU and recS mutations were introduced into rec-deficient strains representative of the alpha (recF), beta (addA5 addB72), gamma (rech342), and epsilon (recG40) epistatic groups.

⇒ **plus difficile** si les unités à annoter sont "noyées" au milieu des autres, en particulier si la segmentation n'est pas évidente.

Discrimination

Plus la proportion de ce qui *doit* être annoté par rapport à ce qui *pourrait* être annoté est faible, plus le poids de la discrimination est élevé :

Définition

$$Discrimination(Flux) = 1 - \frac{|Annotations(Flux)|}{\sum_{i=1}^{nivSeg} |UnitésObtenuesParDécoupage_i(Flux)|}$$

⇒ Nécessité d'une [segmentation de référence](#)

Parties du discours [Marcus et al., 1993] :

I/PRP do/VBP n't/RB feel/VB very/RB ferocious/JJ ./.

$$\text{Discrimination}_{PTB_{POS}} = 0$$

Renommage de gènes [LREC 2012a] :

The yppB:cat and ypbC:cat null alleles rendered cells sensitive to DNA-damaging agents, impaired plasmid transformation (25- and 100-fold), and moderately affected chromosomal transformation when present in an otherwise Rec+ B. subtilis strain. The yppB gene complemented the defect of the recG40 strain. yppB and ypbC and their respective null alleles were termed recU and "recU1" (recU:cat) and recS and "recS1" (recS:cat), respectively. The recU and recS mutations were introduced into rec-deficient strains representative of the alpha (recF), beta (addA5 addB72), gamma (rech342), and epsilon (recG40) epistatic groups.

$$\text{Discrimination}_{\text{Repérage}} = 0,9$$

$$\text{Discrimination}_{\text{Renommage}} = 0,95$$

Délimitation des frontières

Délimiter les frontières consiste à :

- **étendre** ou **rétrécir** l'unité discriminée :
Madame Chirac → *Monsieur et Madame Chirac*

Délimitation des frontières

Délimiter les frontières consiste à :

- **étendre** ou **rétrécir** l'unité discriminée :
Madame Chirac → *Monsieur et Madame Chirac*
- **décomposer** une unité discriminée en plusieurs éléments :
le préfet Érignac → le *préfet* *Érignac*

Délimitation des frontières

Délimiter les frontières consiste à :

- **étendre** ou **rétrécir** l'unité discriminée :
Madame Chirac → *Monsieur et Madame Chirac*
- **décomposer** une unité discriminée en plusieurs éléments :
le préfet Érignac → *le **préfet** **Érignac***
- ou **regrouper** plusieurs unités discriminées en une seule annotation :
Sa Majesté
le roi Mohamed VI → ***Sa Majesté le roi Mohamed VI***

Délimitation

Définition

$$Délimitation(Flux) = \min \left(\frac{Substitutions + Ajouts + Suppressions}{|Annotations(Flux)|}, 1 \right)$$

$$Délimitation_{Repérage} = 0$$

$$Délimitation_{Renommage} = 0$$

$$Délimitation_{PTB_{POS}} = 0$$

$$Délimitation_{EN_{TypesSoustypes}} = 1$$

$$Délimitation_{EN_{Composants}} = 0,3$$

Expressivité du langage d'annotation

Définition

Les degrés d'expressivité du langage d'annotation sont les suivants :

- 0,25 : langages de types
- 0,5 : langages relationnels d'arité 2
- 0,75 : langages relationnels d'arité supérieure à 2
- 1 : langages d'ordre supérieur

$$\begin{aligned} \textit{Expressivité}_{\textit{Repérage}} &= 0,25 \\ \textit{Expressivité}_{\textit{Renommage}} &= 0,25 \end{aligned}$$

Dimension du jeu d'étiquettes

Person			Function		
<i>pers.ind</i> (individual person)	<i>pers.coll</i> (group of persons)		<i>func.ind</i> (individual function)	<i>func.coll</i> (collectivity of functions)	
Location			Production		
<i>administrative</i> (<i>loc.adm.town</i> , <i>loc.adm.reg</i> , <i>loc.adm.nat</i> , <i>loc.adm.sup</i>)	physical (<i>loc.phys.geo</i> , <i>loc.phys.hydro</i> , <i>loc.phys.astro</i>)	facilities (<i>loc.fac</i>), oronyms (<i>loc.oro</i>), address (<i>loc.add.phys</i> , <i>loc.add.elec</i>)	<i>prod.object</i> (manufactured object)	<i>prod.serv</i> (transportation route)	<i>prod.fin</i> (financial products)
			<i>prod.doctr</i> (doctrine)	<i>prod.rule</i> (law)	<i>prod.soft</i> (software)
			<i>prod.art</i>	<i>prod.media</i>	<i>prod.award</i>
Organization			Time		
<i>org.adm</i> (administration)	<i>org.ent</i> (services)		<i>time.date.abs</i> (absolute date), <i>time.date.rel</i> (relative date)	<i>time.hour.abs</i> (absolute hour), <i>time.hour.rel</i> (relative hour)	
Amount					
<i>amount</i> (with unit or general object), including duration					

Types et sous-types utilisés pour l'annotation en EN structurées [LAW 2011]

Dimension du jeu d'étiquettes

Person			Function		
<i>pers.ind</i> (individual person)	<i>pers.coll</i> (group of persons)		<i>func.ind</i> (individual function)	<i>func.coll</i> (collectivity of functions)	
Location			Production		
<i>administrative</i> (<i>loc.adm.town</i> , <i>loc.adm.reg</i> , <i>loc.adm.nat</i> , <i>loc.adm.sup</i>)	<i>physical</i> (<i>loc.phys.geo</i> , <i>loc.phys.hydro</i> , <i>loc.phys.astro</i>)	<i>facilities</i> (<i>loc.fac</i>), oronyms (<i>loc.oro</i>), address (<i>loc.add.phys</i> , <i>loc.add.elec</i>)	<i>prod.object</i> (manufactured object)	<i>prod.serv</i> (transportation route)	<i>prod.fin</i> (financial products)
			<i>prod.doctr</i> (doctrine)	<i>prod.rule</i> (law)	<i>prod.soft</i> (software)
			<i>prod.art</i>	<i>prod.media</i>	<i>prod.award</i>
Organization			Time		
<i>org.adm</i> (administration)	<i>org.ent</i> (services)		<i>time.date.abs</i> (absolute date), <i>time.date.rel</i> (relative date)	<i>time.hour.abs</i> (absolute hour),	<i>time.hour.rel</i> (relative hour)
Amount					
<i>amount</i> (with unit or general object), including duration					

Niveau 1 : *pers*, *func*, *loc*, *prod*, *org*, *time*, *amount* → 7 possibilités (degré de liberté = 6).

Dimension du jeu d'étiquettes

Person			Function		
<i>pers.ind</i> (individual person)	<i>pers.coll</i> (group of persons)		<i>func.ind</i> (individual function)	<i>func.coll</i> (collectivity of functions)	
Location			Production		
<i>administrative</i> (<i>loc.adm.town</i> , <i>loc.adm.reg</i> , <i>loc.adm.nat</i> , <i>loc.adm.sup</i>)	<i>physical</i> (<i>loc.phys.geo</i> , <i>loc.phys.hydro</i> , <i>loc.phys.astro</i>)	<i>facilities</i> (<i>loc.fac</i>), oronyms (<i>loc.oro</i>), address (<i>loc.add.phys</i> , <i>loc.add.elec</i>)	<i>prod.object</i> (manufactured object)	<i>prod.serv</i> (transportation route)	<i>prod.fin</i> (financial products)
			<i>prod.doctr</i> (doctrine)	<i>prod.rule</i> (law)	<i>prod.soft</i> (software)
			<i>prod.art</i>	<i>prod.media</i>	<i>prod.award</i>
Organization			Time		
<i>org.adm</i> (administration)	<i>org.ent</i> (services)		<i>time.date.abs</i> (absolute date), <i>time.date.rel</i> (relative date)	<i>time.hour.abs</i> (absolute hour), <i>time.hour.rel</i> (relative hour)	
Amount					
<i>amount</i> (with unit or general object), including duration					

Niveau 1 : *pers*, *func*, *loc*, *prod*, *org*, *time*, *amount* → 7 possibilités (degré de liberté = 6).

Niveau 2 : *prod.object*, *prod.serv*, *prod.fin*, *prod.soft*, *prod.doctr*, *prod.rule*, *prod.art*, *prod.media*, *prod.award* → 9 possibilités (degré de liberté = 8).

Dimension du jeu d'étiquettes

Person			Function		
<i>pers.ind</i> (individual person)	<i>pers.coll</i> (group of persons)		<i>func.ind</i> (individual function)	<i>func.coll</i> (collectivity of functions)	
Location			Production		
<i>administrative</i> <i>(loc.adm.town,</i> <i>loc.adm.reg,</i> <i>loc.adm.nat,</i> <i>loc.adm.sup)</i>	<i>physical</i> <i>(loc.phys.geo,</i> <i>loc.phys.hydro,</i> <i>loc.phys.astro)</i>	<i>facilities</i> <i>(loc.fac),</i> <i>oronyms</i> <i>(loc.oro),</i> <i>address</i> <i>(loc.add.phys,</i> <i>loc.add.elec)</i>	<i>prod.object</i> (manufactured object)	<i>prod.serv</i> (transportation route)	<i>prod.fin</i> (financial products)
			<i>prod.doctr</i> (doctrine)	<i>prod.rule</i> (law)	<i>prod.soft</i> (software)
			<i>prod.art</i>	<i>prod.media</i>	<i>prod.award</i>
Organization			Time		
<i>org.adm</i> (administration)	<i>org.ent</i> (services)		<i>time.date.abs</i> (absolute date), <i>time.date.rel</i> (relative date)	<i>time.hour.abs</i> (absolute hour), <i>time.hour.rel</i> (relative hour)	
Amount					
<i>amount</i> (with unit or general object), including duration					

Niveau 1 : *pers, func, loc, prod, org, time, amount* → 7 possibilités (degré de liberté = 6).

Niveau 2 : *prod.object, prod.serv, prod.fin, prod.soft, prod.doctr, prod.rule, prod.art, prod.media, prod.award* → 9 possibilités (degré de liberté = 8).

Niveau 3 : *loc.adm.town, loc.adm.reg, loc.adm.nat, loc.adm.sup* → 4 possibilités (degré de liberté = 3).

Dimension du jeu d'étiquettes

Degré de liberté

$$\nu = \nu_1 + \nu_2 + \dots + \nu_m$$

où ν_i est le degré de liberté maximal que l'annotateur a dans le choix de la i^{eme} sous-étiquette ($\nu_i = n_i - 1$).

Dimension du jeu d'étiquettes

$$\text{Dimension}(\text{Flux}) = \min\left(\frac{\nu}{\tau}, 1\right)$$

où τ est le seuil à partir duquel on considère le jeu d'étiquettes comme arbitrairement grand (déterminé expérimentalement).

$$\text{Dimension}_{\text{Repérage}} = 0$$

$$\text{Dimension}_{\text{Renommage}} = 0,04$$

$$\text{Dimension}_{\text{EN}_{\text{TypesSoustypes}}} = 0,34$$

Degré d'ambiguïté : ambiguïté résiduelle

Utiliser les traces laissées par les annotateurs :



[...] *<EukVirus>3CDproM</EukVirus> can process both structural and nonstructural precursors of the <EukVirus uncertainty-type = "too-generic"><taxon>poliovirus</taxon> polyprotein</EukVirus> [...].*

Définition

$$Ambiguïté_{Res}(Flux) = \frac{|Annotations_{amb}|}{|Annotations|}$$

$$Ambiguïté_{ResRepérage} = 0,04$$

$$Ambiguïté_{ResRenommage} = 0,02$$

Degré d'ambiguïté : ambiguïté théorique

Proportion des unités à annoter qui correspond à des vocables ambigus.

Définition

$$Ambiguïté_{Th}(Flux) = \frac{\sum_{voc_i=1}^{|Voc(Flux)|} (Ambig(voc_i) * freq(voc_i, Flux))}{|Unités(Flux)|}$$

avec

$$Ambig(voc_i) = \begin{cases} 1 & \text{si } |Étiquettes(voc_i)| > 1 \\ 0 & \text{sinon} \end{cases}$$

$$Ambiguïté_{Th_{Repérage}} = 0,01$$

→ Ne s'applique pas aux relations de renommage (TAE2).

Poids du contexte

- **taille de la fenêtre** de signal source à prendre en compte :

- ▶ La phrase :

I/PRP do/VBP n't/RB feel/VB very/RB ferocious/JJ ./.

- ▶ ... ou plus :

Fabien Lévêque : C'est bien fait , avec **Gouffran** maintenant . **Gouffran** qui va tenter sa chance , et ça fait le but . Le but !

Xavier Gravelaine : Oh la la la la !

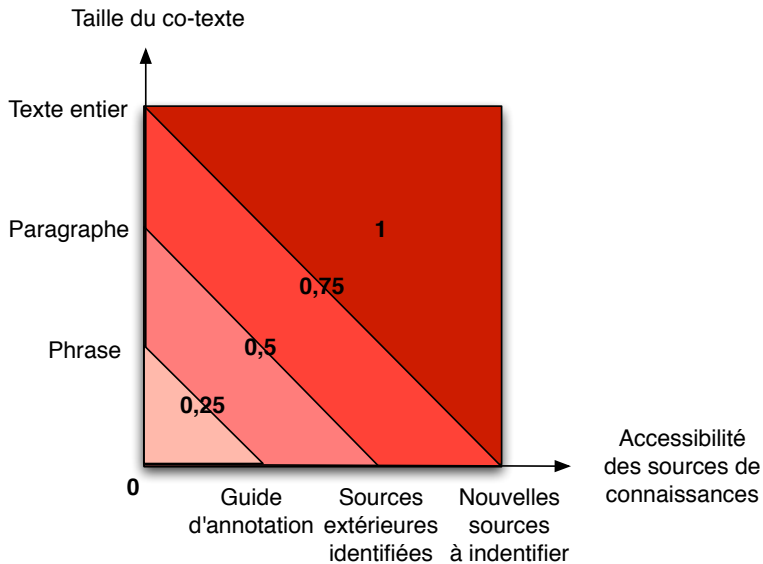
Fabien Lévêque : Et le but du plus breton des **Girondins** . C'est **Yoann Gourcuif** qui vient mettre un quatrième but ici au **stade de France** . Le cauchemar continue pour le **VOC** . Quatre à zéro en faveur des **Girondins** .

id=518

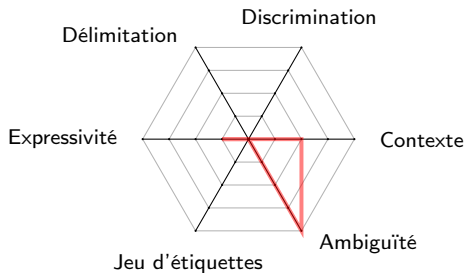
- nombre de **connaissances** à mobiliser ou degré d'accessibilité des sources de connaissances qui sont consultées :

- ▶ guide d'annotation
 - ▶ nomenclatures (Swiss-Prot)
 - ▶ nouvelles sources à trouver (Wikipedia, etc.)

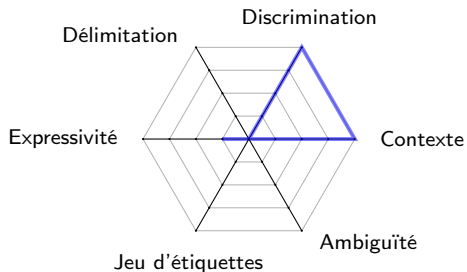
Poids du contexte



Synthèse des dimensions de complexité

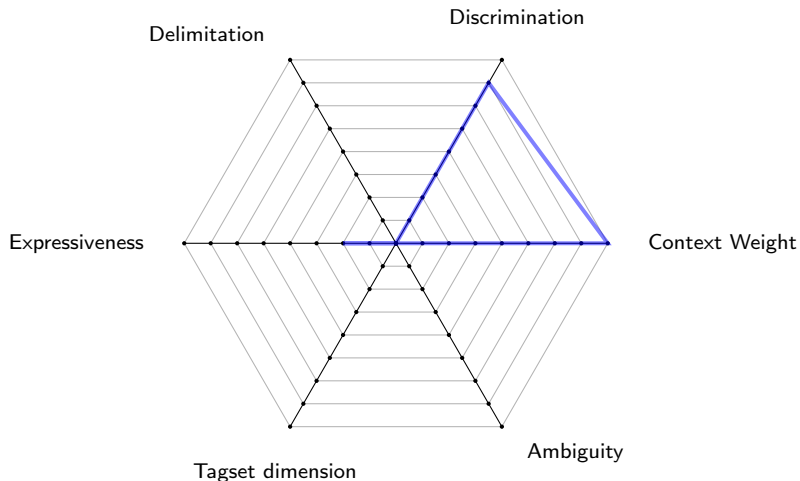


Classification des pronoms
(*it* anaphoriques ou impersonnels)



Repérage des noms de gènes

Synthèse des dimensions de complexité



Campagne de renommage de gènes (2 TAE)

- 1 Introduction
- 2 Une maîtrise insuffisante du processus d'annotation
- 3 Organiser une campagne d'annotation
- 4 Analyser la complexité d'une campagne d'annotation
- 5 Évaluer les annotations manuelles résultantes**
 - Mesurer la qualité ?
 - Synthétiser les données
 - Choisir des mesures appropriées
 - Donner du sens
- 6 Conclusion et perspectives

Qu'est-ce que la qualité d'une annotation ?

Processus d'**interprétation** \Rightarrow pas de "vérité absolue" \Rightarrow impossible d'évaluer la validité d'une annotation

Mais il est possible de mesurer sa **fiabilité** : la cohérence avec laquelle les annotateurs ont réalisé l'annotation

[Gut and Bayerl, 2004]

- l'**accord inter-annotateur** permet de mesurer la **stabilité** de l'annotation
- l'**accord intra-annotateur** (de l'annotateur avec lui-même, plus tard dans la campagne), donne une indication de la **reproductibilité** de l'annotation

Mesures utilisées aujourd'hui [Artstein and Poesio, 2008]

Accord observé :

- + simple à calculer
 - nécessite une quantification du "cas négatif" (ce qui n'est pas annoté, mais pourrait l'être : les annotables)
 - ne prend pas en compte le hasard
- ⇒ très influençable et inutilisable dans de nombreux cas

Mesures utilisées aujourd'hui [Artstein and Poesio, 2008]

Accord observé :

- + simple à calculer
- nécessite une quantification du "cas négatif" (ce qui n'est pas annoté, mais pourrait l'être : les annotables)
- ne prend pas en compte le hasard
- ⇒ très influençable et inutilisable dans de nombreux cas

Famille des κ (κ [Cohen, 1960], π [Scott, 1955, Carletta, 1996]) :

- + prennent en compte le hasard
- nécessitent une quantification des annotables
- ⇒ inutilisables dans de nombreux cas

Mesures utilisées aujourd'hui [Artstein and Poesio, 2008]

Accord observé :

- + simple à calculer
- nécessite une quantification du "cas négatif" (ce qui n'est pas annoté, mais pourrait l'être : les annotables)
- ne prend pas en compte le hasard
- ⇒ très influençable et inutilisable dans de nombreux cas

Famille des κ (κ [Cohen, 1960], π [Scott, 1955, Carletta, 1996]) :

- + prennent en compte le hasard
- nécessitent une quantification des annotables
- ⇒ inutilisables dans de nombreux cas

F-mesure :

- + pas de quantification des annotables [Hripcsak and Rothschild, 2005]
- ne prend pas en compte le hasard
- ⇒ trop favorable

Synthétiser les données : matrice de confusion

Matrice de confusion de la campagne d'annotation de relations de renommage de gènes :

		A1			
		Former	New	Rien	Total
A2	Former	71	13	23	107
	New	8	69	15	92
	Rien	7	8	18 840	18 855
	Total	86	90	18 878	19 054

[Hripcsak and Heitjan, 2002] :

showing the two-by-two contingency table with its marginal totals is probably as informative as any measure.

Synthétiser les données : matrice de confusion

Matrice de confusion de la campagne d'annotation de relations de renommage de gènes :

		A1			
		Former	New	Rien	Total
A2	Former	71	13	23	107
	New	8	69	15	92
	Rien	7	8	18 840	18 855
	Total	86	90	18 878	19 054

[Hripcsak and Heitjan, 2002] :

showing the two-by-two contingency table with its marginal totals is probably as informative as any measure.

Mais n'est utilisable que pour **deux** annotateurs

Synthétiser les données : tableau de Krippendorff

[Krippendorff, 2004]

	A1	A2	A3
token 1	a	a	c
token 2	b	b	a
token 3	b	a	a
token 4	a	a	c
token 5	a	b	c

Synthétiser les données : tableau de Krippendorff

[Krippendorff, 2004]

	A1	A2	A3
token 1	a	a	c
token 2	b	b	a
token 3	b	a	a
token 4	a	a	c
token 5	a	b	c

Mais n'est utilisable que pour des **petits** corpus

Synthétiser les données : similarités entre catégories

[LREC 2012a]

Probabilités (de type $P(C_2|C_1)$) qu'un annotateur affecte un élément à une catégorie (C_2) sachant qu'un autre l'affecte à une autre catégorie (C_1) :

$$P(C_2|C_1) = \frac{n_{1C_1,2C_2} + n_{2C_1,1C_2}}{n_{C_1}}$$

Puis :

$$Sim(C_1, C_2) = \frac{P(C_2|C_1) + P(C_1|C_2)}{2}$$

Similarités entre catégories pour la campagne renommage de gènes [LREC 2012a]

	<i>Sim</i>
<i>Sim</i> (Former, New)	0,112096
<i>Sim</i> (Former, Rien)	0,078117
<i>Sim</i> (New, Rien)	0,063491

⇒ *Former* et *New* sont plus ambiguës entre elles qu'avec *Rien*

Critères généraux à prendre en compte (groupe de travail)

- **Couverture** de l'annotation (discrimination):

- ▶ pavage partiel (renommage) :

*The yppB gene complemented the defect of the recG40 strain.
yppB and ypbC and their respective null alleles were termed recU
and "recU1" (recU:cat) and recS and "recS1" (recS:cat),
respectively.*

- ▶ pavage complet (POS) :

I/PRP do/VBP n't/RB feel/VB very/RB ferocious /JJ./.

- **Homogénéité** de la segmentation (délimitation) :

- ▶ homogène (POS, renommage)
- ▶ hétérogène (EN, football) :

le préfet Érignac

Sa Majesté le roi Mohamed VI

Critères transversaux à prendre en compte

Le **jeu d'étiquettes** (dimension du jeu d'étiquettes) :

- simple : POS pour le Penn Treebank
- structuré : EN structurées → certaines étiquettes sont plus “proches” que d'autres (erreur moins grave)

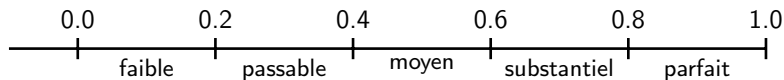
L'influence de la **prévalence** [LREC 2012a] → prise en compte du hasard indispensable.

Cas de figure pour les unités

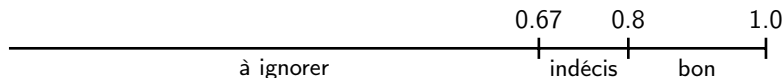
- 1 Pavage complet avec segmentation homogène (POS) :
 - type κ simple ou pondéré (annotables évidents)
- 2 Pavage partiel avec segmentation homogène (*it*) :
 - ajouter une catégorie “vide” pour se ramener au cas précédent ⇒ effet de **prévalence** possible.
- 3 Pavage complet avec segmentation hétérogène (syntaxe en constituants) :
 - type κ pondéré (si structure), mais question de la détermination des **poids** ou des distances
- 4 Pavage partiel avec segmentation hétérogène (football, EN) :
 - type κ simple ou pondéré, mais les **annotables** sont à définir [LREC 2012]

Échelles d'interprétation des Kappas

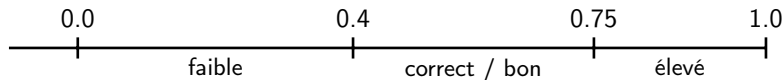
[Landis and Koch, 1977]



[Krippendorff, 1980]



[Green, 1997]



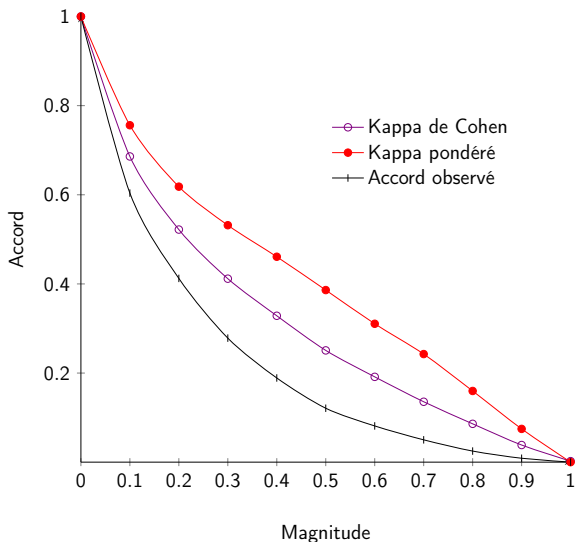
Donner du sens au résultat obtenu [COLING 2012a]

Création d'un outil "Richter" qui :

- prend en entrée une annotation de référence (réelle ou générée automatiquement)
- génère des annotations dégradées en fonction d'une **magnitude** (de 0 à 1)
- applique une ou des mesures d'accord inter-annotateurs sur chaque ensemble d'annotations (correspondant à une magnitude)

Sur le corpus TCOF-POS [TALN 2012]

Pas de prévalence, mais proximité entre catégories prise en compte :



- 1 Introduction
- 2 Une maîtrise insuffisante du processus d'annotation
- 3 Organiser une campagne d'annotation
- 4 Analyser la complexité d'une campagne d'annotation
- 5 Évaluer les annotations manuelles résultantes
- 6 Conclusion et perspectives**

Annotation manuelle de corpus et TAL

→ **Vision unifiée** de l'annotation manuelle de corpus pour le TAL, construite à partir de :

- expériences des campagnes d'annotation (participation ou gestion)
- collaborations multiples
- connaissance d'autres campagnes d'annotation

→ **Méthodologie** détaillée :

- en amont : dimensions de complexité
- au centre : évaluation
- déroulement : identification des processus en œuvre et des acteurs

→ **Identification** des biais et des apports réels des outils de pré-annotation en morpho-syntaxe (non présenté ici) [LAW 2010]





Travaux en cours

- Groupe de travail sur les accords inter-annotateurs : application de l'outil Richter sur **corpus réels, relations**, détermination de la **prévalence**
- Annotation manuelle d'un corpus arboré du français librement disponible [Candito and Seddah, 2012] : correction d'annotations automatiques à l'aide d'un **jeu** (GWAP)
- Charte **Éthique** (ATALA, AFCP, APROGED, Cap Digital, ELRA/ELDA)

En perspective

- ① Développer un module générique de calcul des dimensions de complexité et l'intégrer dans un ou des outils d'annotation
- ② Étendre les expériences sur l'annotation assistée :
 - ▶ établir une typologie fine des méthodes d'annotation assistée
 - ▶ mettre au point un protocole d'expérience pour tester l'influence de la pré-annotation sur différentes tâches d'annotation
 - ▶ comparer pré-annotation et *active learning*
- ③ Articuler complexité et évaluation

Merci de votre attention

-  Alphonse, E., Aubin, S., Bessières, P., Bisson, G., Hamon, T., Laguarigue, S., Nazarenko, A., Manine, A.-P., Nédellec, C., Vetah, M. O. A., Poibeau, T., and Weissenbacher, D. (2004).
Event-based information extraction for the biomedical the CADERIGE project.
In Proceedings of the JNLPBA COLING 2004 Workshop, Genève, Suisse.
-  Artstein, R. and Poesio, M. (2008).
Inter-coder agreement for computational linguistics.
Computational Linguistics, 34(4):555–596.
-  Benzitoun, C., Fort, K., and Sagot, B. (2012).
TCOF-POS : un corpus libre de français parlé annoté en morphosyntaxe.
In Actes de Traitement Automatique des Langues Naturelles (TALN), pages 99–112, Grenoble, France.
-  Böhmová, A., Hajič, J., Hajičová, E., and Hladká, B. (2001).
The prague dependency treebank: Three-level annotation scenario.

In Abeillé, A., editor, *Treebanks: Building and Using Syntactically Annotated Corpora*. Kluwer Academic Publishers.



Bonneau-Maynard, H., Rosset, S., Ayache, C., Kuhn, A., and Mostefa, D. (2005).

Semantic annotation of the French Media dialog corpus.
In *Proceedings of the InterSpeech*, Lisbonne, Portugal.



Candito, M. and Seddah, D. (2012).

Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical.

In *Actes de Traitement Automatique des Langues Naturelles (TALN)*, Grenoble, France.



Carletta, J. (1996).

Assessing agreement on classification tasks: the kappa statistic.
Computational Linguistics, 22:249–254.



Carmen, M., Felt, P., Haertel, R., Lonsdale, D., McClanahan, P., Merkling, O., Ringger, E., and Seppi, K. (2010).

Tag dictionaries accelerate manual annotation.

In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, La Valette, Malte. European Language Resources Association (ELRA).



Chamberlain, J., Poesio, M., and Kruschwitz, U. (2008).

Addressing the resource bottleneck to create large-scale annotated texts.

In *STEP '08: Proceedings of the 2008 Conference on Semantics in Text Processing*, pages 375–380, Morristown, NJ, USA. Association for Computational Linguistics.



Cohen, J. (1960).

A coefficient of agreement for nominal scales.

Educational and Psychological Measurement, 20(1):37–46.



Cunningham, H., Maynard, D., Bontcheva, K., and Tablan, V. (2002).

GATE: a framework and graphical development environment for robust NLP tools and applications.

In *Proceedings of the 40th Anniversary Meeting of the Association for Computational Linguistics*, Philadelphia, USA.



de la Clergerie, E. V. (2008).

A collaborative infrastructure for handling syntactic annotations.

In *Proceedings of the First International Workshop on Automated Syntactic Annotations for interoperable Language Resources*, Hong-Kong, Chine.



Felt, P., Merklings, O., Carmen, M., Ringger, E., Lemmon, W., Seppi, K., and Haertel, R. (2010).

CCASH: A web application framework for efficient, distributed language resource development.

In Chair), N. C. C., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, La Valette, Malte. European Language Resources Association (ELRA).



Fort, K., Adda, G., and Cohen, K. B. (2011a).

Amazon Mechanical Turk: Gold mine or coal mine?
Computational Linguistics (editorial), 37(2):413–420.



Fort, K. and Claveau, V. (2012).

Annotating football matches: Influence of the source medium on manual annotation.

In Proceedings of the International Conference on Language Resources and Evaluation (LREC), Istanbul, Turquie.

6 pages.



Fort, K., Ehrmann, M., and Nazarenko, A. (2009).

Vers une méthodologie d'annotation des entités nommées en corpus ?
In Actes de Traitement Automatique des Langues Naturelles (TALN), Senlis, France.

11 pages.



Fort, K., François, C., Galibert, O., and Ghribi, M. (2012a).

Analyzing the impact of prevalence on the evaluation of a manual annotation campaign.

In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Istanbul, Turquie.

7 pages.



Fort, K., Nazarenko, A., and Ris, C. (2011b).

Corpus linguistics for the annotation manager.

In *Proceedings of the Corpus Linguistics Conference*, Birmingham, Angleterre.

13 pages.



Fort, K., Nazarenko, A., and Rosset, S. (2012b).

Modeling the complexity of manual annotation tasks: a grid of analysis.





In *Proceedings of the International Conference on Computational Linguistics (COLING)*, Mumbai, Inde.



Fort, K. and Sagot, B. (2010).

Influence of pre-annotation on POS-tagged corpus development.

In *Proceedings of the Fourth ACL Linguistic Annotation Workshop*, pages 56–63, Uppsala, Suède.

-  Green, A. M. (1997).
Kappa statistics for multiple raters using categorical classifications.
In *Proceedings of the Twenty-Second Annual Conference of SAS Users Group*, San Diego, USA.
-  Grouin, C., Rosset, S., Zweigenbaum, P., Fort, K., Galibert, O., and Quintard, L. (2011).
Proposal for an extension of traditional named entities: From guidelines to evaluation, an overview.
In *Proceedings of the 5th Linguistic Annotation Workshop*, pages 92–100, Portland, Oregon, USA.
Poster.
-  Gut, U. and Bayerl, P. S. (2004).
Measuring the reliability of manual annotations of speech corpora.
In *Proceedings of the Speech Prosody*, pages 565–568, Nara, Japan.
-  Habert, B. (2000).

Corpus. Méthodologie et applications linguistiques, chapter
Détournements d'annotation : armer la main et le regard, pages
106–120.

Champion and Presses Universitaires de Perpignan.



Habert, B. (2005).

Portrait de linguiste(s) à l'instrument.

Texto!, vol. X(4).



Hripcsak, G. and Heitjan, D. F. (2002).

Measuring agreement in medical informatics reliability studies.

Journal of Biomedical Informatics, 35(2):99–110.



Hripcsak, G. and Rothschild, A. S. (2005).

Agreement, the f measure, and reliability in information retrieval.

Journal of the American Medical Informatics Association (JAMIA),
12(3):296–298.



Jourde, J., Manine, A.-P., Veber, P., Fort, K., Bossy, R., Alphonse, E.,
and Bessières, P. (2011).

BioNLP shared task 2011 – bacteria gene interactions and renaming.
In *Proceedings of the BioNLP Shared Task 2011 Workshop*, pages 65–73, Portland, Oregon, USA. Association for Computational Linguistics.

 Kaplan, D., Iida, R., and Tokunaga, T. (2010).

Annotation process management revisited.

In *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, pages 365 – 366, La Valette, Malte.

 Kim, J.-D., Ohta, T., and Tsujii, J. (2008).

Corpus annotation for mining biomedical events from literature.

BMC Bioinformatics, 9(1):10.

 Krippendorff, K. (1980).

Content Analysis: An Introduction to Its Methodology, chapter 12.

Sage, Beverly Hills, CA., USA.

 Krippendorff, K. (2004).

Content Analysis: An Introduction to Its Methodology, second edition, chapter 11.

Sage, Thousand Oaks, CA., USA.



Landis, J. R. and Koch, G. G. (1977).

The measurement of observer agreement for categorical data.

Biometrics, 33(1):159–174.



Landragin, F., Poibeau, T., and Victorri, B. (2012).

ANALEC: a new tool for the dynamic annotation of textual data.





In Chair), N. C. C., Choukri, K., Declerck, T., Doğan, M. U., Maegaard, B., Mariani, J., Odijk, J., and Piperidis, S., editors, *Proceedings of the International Conference on Language Resources and Evaluation (LREC)*, Istanbul, Turquie. European Language Resources Association (ELRA).



Leech, G. (1997).

Corpus annotation: Linguistic information from computer text corpora, chapter Introducing corpus annotation, pages 1–18.

Longman, Londres, Angleterre.

-  Leech, G. (2005).
Developing Linguistic Corpora: a Guide to Good Practice, chapter
Adding Linguistic Annotation, pages 17–29.
Oxford: Oxbow Books.
-  Marcus, M., Santorini, B., and Marcinkiewicz, M. A. (1993).
Building a large annotated corpus of English : The Penn Treebank.
Computational Linguistics, 19(2):313–330.
-  Mathet, Y., Widlöcher, A., Fort, K., François, C., Galibert, O., Grouin, C., Kahn, J., Rosset, S., and Zweigenbaum, P. (2012).
Manual corpus annotation: Evaluating the evaluation metrics.
In Proceedings of the International Conference on Computational Linguistics (COLING), Mumbai, Inde.
Poster.
-  Rosset, S., Grouin, C., Fort, K., Galibert, O., Kahn, J., and Zweigenbaum, P. (2012).
Structured named entities in two distinct press corpora: Contemporary broadcast news and old newspapers.

In *Proceedings of the 6th Linguistic Annotation Workshop (LAW VI)*, pages 40–48, Jeju, République de Corée.



Scott, W. A. (1955).

Reliability of content analysis: The case of nominal scale coding.

Public Opinion Quarterly, 19(3):321–325.



von Ahn, L. (2006).

Games with a purpose.

IEEE Computer Magazine, pages 96–98.



Voormann, H. and Gut, U. (2008).

Agile corpus creation.

Corpus Linguistics and Linguistic Theory, 4(2):235–251.



Widlöcher, A. and Mathet, Y. (2009).

La plate-forme Glozz : environnement d'annotation et d'exploration de corpus.

In *Actes de Traitement Automatique des Langues Naturelles (TALN)*, Senlis, France.

Maha Ghribi
Simonetta Vietri Annibale Elia
Claire Ris Michel Guoet Maud Ehrmann
Gilles Adda Benoît Sagot Robert Bossy Massimo Poesio
Sophie Rosset Guy Perrier Adeline Nazarenko
Philippe Veber Geoffroy Peeters Jennifer Planul
Emilio D'Agostino Kevin Bretonnel Cohen
Joseph Mariani Dominique Besagni Veronika Lux-Pogodalla
Jonathan Marchand Jonathan Chamberlain Géraldine Walther
Mathieu Lafourcade Christophe Benzitoun Cyril Grouin
Claire François Philippe Bessières Pierre Zweigenbaum
Udo Kruschwitz Joseph Le Roux Antoine Widlöcher
Bernard Lang Erick Alphonse Alain-Pierre Manine
Vincent Claveau Ludovic Quintard Fabienne Venant
Olivier Galibert Bruno Guillaume Yann Mathet
Juliette Kahn
Julien Jourde