

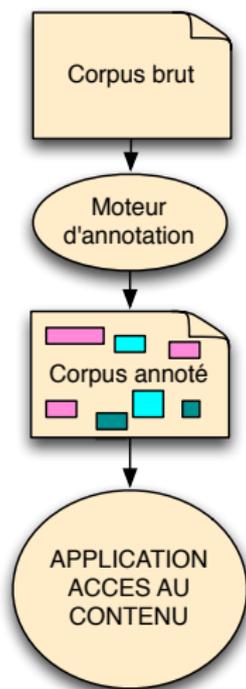
(Em)portés par la foule ?
comprendre et contrôler le *crowdsourcing* pour
créer des données (annotées) de qualité

Karën Fort

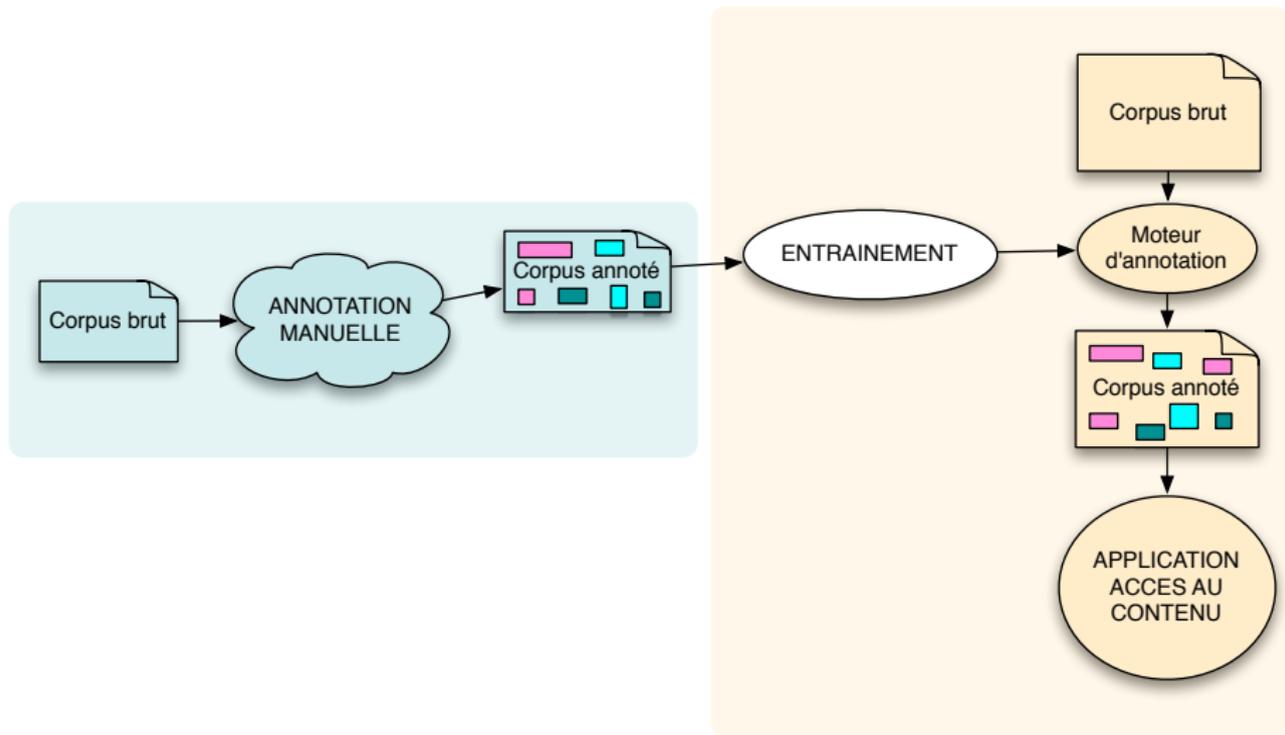
10 mars 2014



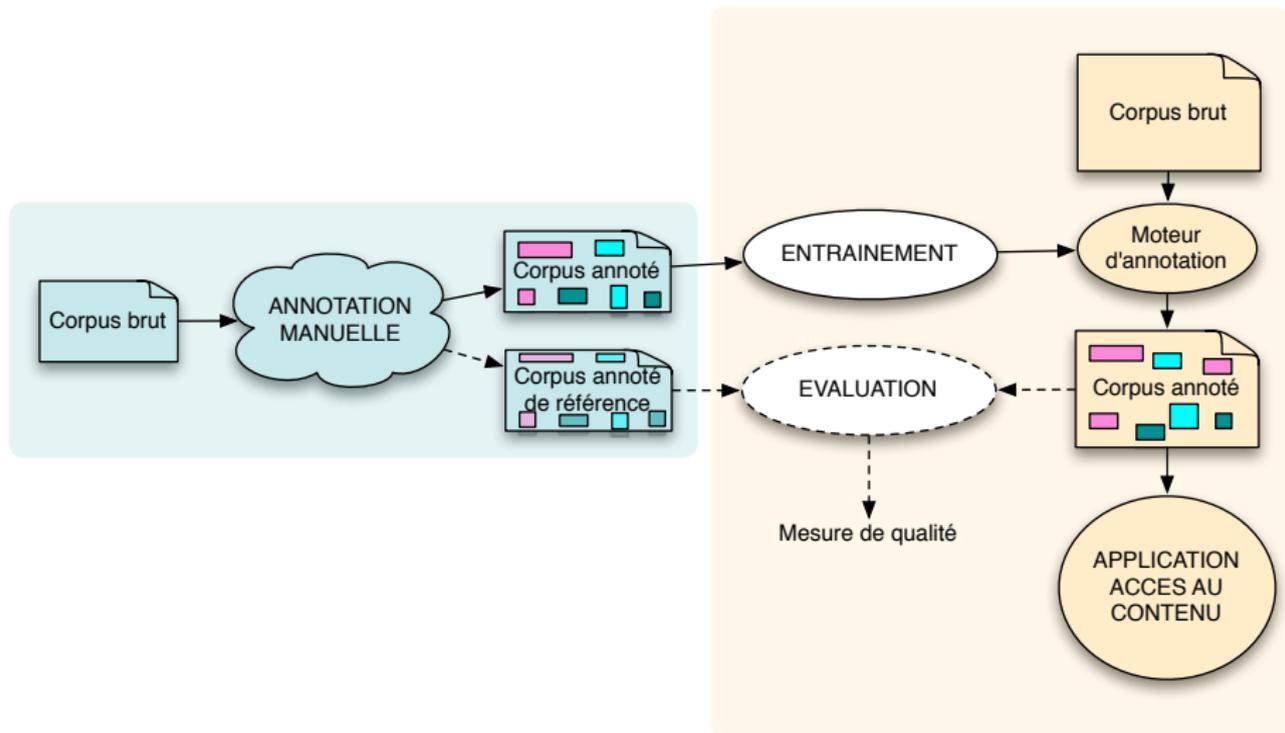
Corpus annotés et Traitement Automatique des Langues



Corpus annotés et Traitement Automatique des Langues



Corpus annotés et Traitement Automatique des Langues



Des ressources diverses, de complexité croissante

Diversité accrue :

- des **médias** : texte, parole, musique, vidéo
- des **champs d'applications** : linguistique, domaines spécialisés

Complexité accrue :

- 1960-1990 : morpho-syntaxe
- 1990-2000 : morpho-syntaxe et syntaxe (corpus arborés), entités nommées simples, « sens » (WordNet, FrameNet)
- Depuis 2000 : annotations sémantiques variées (opinions, émotions, etc.), discours, entités nommées structurées, etc.

Taille accrue :

- 1961 : Brown Corpus, 1 million de mots
- 1995 : British National Corpus (BNC), 100 millions de mots

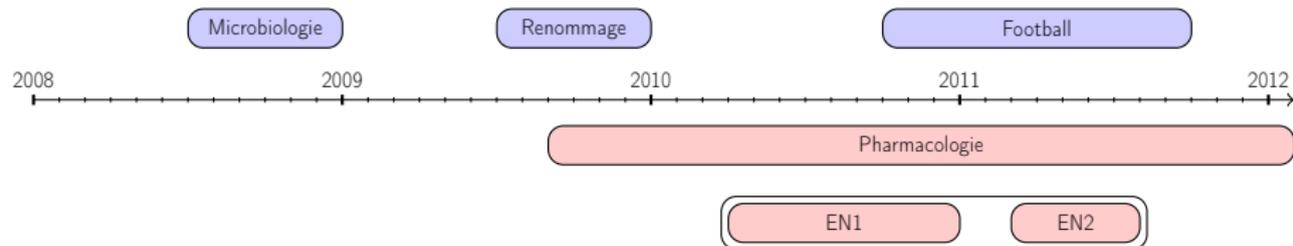
Le coût élevé (prohibitif?) de la création de ressources

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

- 1,8 millions de mots annotés en morpho-syntaxe et syntaxe

⇒ 5 ans, 22 personnes (max. 17 en parallèle), 600 000 dollars

Quæro : des corpus annotés pour 10 millions d'euros, dont...



Quæro : programme collaboratif franco-allemand public-privé portant sur le traitement automatique de contenus numériques multimédias multilingues

Microbiologie : noms d'espèces, de gènes et de protéines

Renommage : relations de renommage de gènes

Football : entités nommées, actions et relations

Pharmacologie : entités nommées, termes et relations en pharmacologie

EN 1 & 2 : entités nommées structurées

Participation à des campagnes

Gestion de campagnes

1 Introduction

2 Myriadiser pour créer des ressources

- Profiter des connaissances de la foule
- Profiter des connaissances scolaires de la foule
- Profiter des capacités d'apprentissage de la foule
- Penser la myriadisation

3 Évaluer la qualité de l'annotation manuelle

4 Gérer la complexité de l'annotation manuelle

5 Conclusion

Quelques exemples de myriadisations

Wikipedia, Projet Gutenberg :

- bénévole (non rémunérée)
- directe (le but à atteindre n'est pas masqué)

Games With A Purpose (Jeux ayant un but ou GWAP) :

- bénévole (non rémunérée)
- indirecte (le but à atteindre est plus ou moins masqué)

Amazon Mechanical Turk :

- rémunérée
- directe

Voir [Geiger et al., 2011] pour un état de l'art détaillé des taxinomies de la myriadisation

Quelques exemples de myriadisations

Wikipedia, Projet Gutenberg :

- bénévole (non rémunérée)
- directe (le but à atteindre n'est pas masqué)

Games With A Purpose (Jeux ayant un but ou GWAP) :

- bénévole (non rémunérée)
- indirecte (le but à atteindre est plus ou moins masqué)

Amazon Mechanical Turk :

- rémunérée
- directe

Voir [Geiger et al., 2011] pour un état de l'art détaillé des taxinomies de la myriadisation

JeuxDeMots : jouer à associer des idées. . .

. . . pour créer un réseau lexical [Lafourcade and Joubert, 2008]

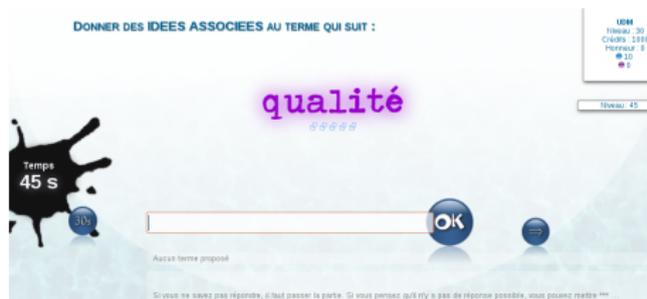
Plus de **6 millions de relations** créées
et **constamment** mises à jour

Mécanisme très élaboré :

- jeu par paires
- défis entre joueurs
- procès, etc.

Évaluation difficile :

- pas de référence
- ⇒ création d'une application pour valider la ressource
[Lafourcade et al., 2011]

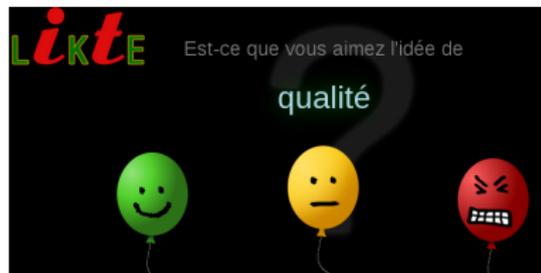


JeuxDeMots : jouer à associer des idées...

... pour créer un réseau lexical [Lafourcade and Joubert, 2008]

Galaxie de **mini-jeux** périphériques :

- LikeIt : 25 000 termes polarisés en 1 mois, 150 000 votes
- à comparer avec Polarimots : 7 473 mots polarisés, 3 annotateurs [Gala and Brun, 2012]
- également SexIt, ColorIt, PolitIt...



LREC 2014

Propa-L : a semantic filtering service from a lexical network created using Games With A Purpose. M. Lafourcade et **K. Fort**

Phrase Detectives : jouer au détective...

... pour annoter des anaphores [Chamberlain et al., 2008]

Corpus annoté de 200 000 mots :

- corpus pré-annoté
- instructions détaillées
- phase de formation
- 2 manières de jouer

Évaluation :

- corpus de référence
- accord inter-annotateurs élevé [Chamberlain et al., 2009]

Échec : identification de propriétés
Jon, the postman, delivered the letter

DETECTIVES CONFERENCE
Another detective has made a decision about a phrase, other than it refers to another phrase, it has not been mentioned before, it is a property or it does not refer to anything. Do you agree with them?

SEARCH CLUES
What's the **blue** or **orange** doing in the text? Is it referring to something else in the text? Try to find the closest mention of the phrase.
What's the **blue** or **orange** doing with it to move the one thing in the text to the closest mention of the phrase to make relations agreement points.
Always look for the **closest** previous mention of the phrase to make relations agreement points.

Feedback

DETECTIVES CONFERENCE
The phrase in blue is the **closest** phrase that refers to the phrase in orange.

Disagree Agree

NAME THE CULPRIT
Has the phrase shown in **orange** been mentioned before in the text or is it a property? Use your mouse to select the **closest phrase(s)** if it has been mentioned before.

SEARCH CLUES
Phrases beginning with a, an or the can be used to identify an object in the text. For example: "The postman delivered a letter" or "John owns a laptop".
1. As an object They can be used to identify something about an object. For example: "The postman delivered a letter" describes the object "letter" or "John owns a laptop".
2. As a property They can be used to identify something about an object. For example: "The postman delivered a letter" describes the object "letter" or "John owns a laptop".
If you think the phrase describes a property try to select the **closest phrase** it refers to.

Not mentioned before This is a property Done

Feedback

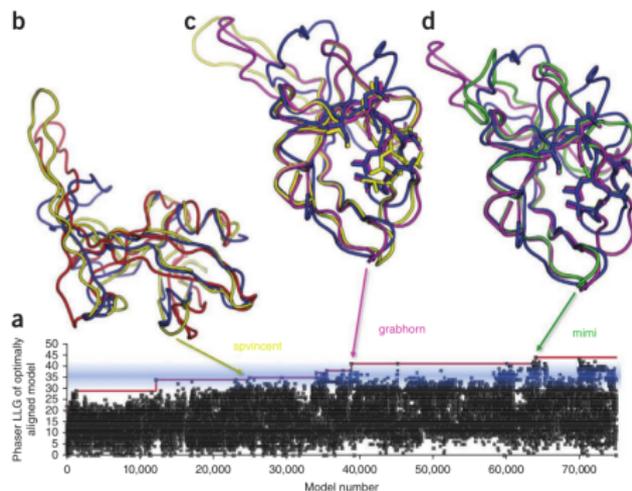
FoldIt : jouer à replier des protéines. . .

. . . pour résoudre des problèmes scientifiques [Khatib et al., 2011]

Résolution de la structure cristalline de la protéine responsable de la propagation du virus du SIDA chez les macaques rhésus

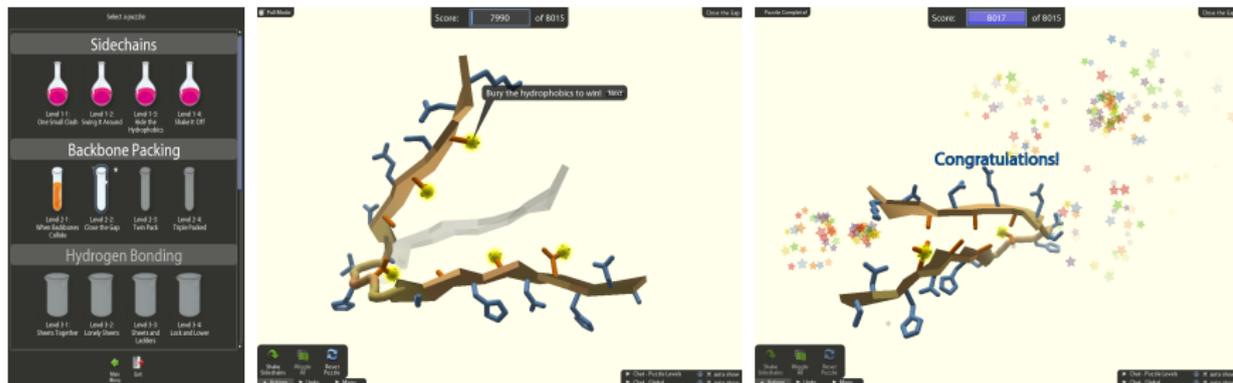
Solution à un problème non résolu depuis une dizaine d'années :

- trouvée en quelques semaines
- par étapes
- par une équipe de joueurs
- qui permettra la création d'antirétroviraux



FoldIt : jouer à replier des protéines. . .

. . . sans connaissances préalables en biochimie [Cooper et al., 2010]



Formation par étapes :

- tutoriel décomposé par concepts
- ensemble de puzzles pour chaque concept
- accès aux puzzles suivants si le nombre de points gagnés est suffisant

Mettre au jour la partie immergée de l'iceberg

Une solution prometteuse :

- massive et rapide
- à coût réduit
- pour des productions variées. . .
- . . . et dynamiques



Mettre au jour la partie immergée de l'iceberg

Une solution prometteuse :

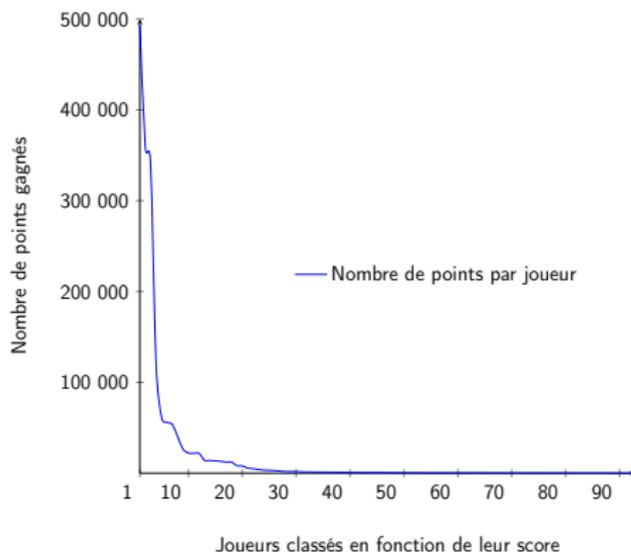
- massive et rapide
- à coût réduit
- pour des productions variées. . .
- . . . et dynamiques

Un phénomène encore peu étudié :

- déconstruire les légendes
 - ▶ Amazon Mechanical Turk
 - ▶ GWAP
- évaluer les ressources produites
- identifier et outiller la complexité



Myriadisation : une foule de « non-experts » ? (GWAP)

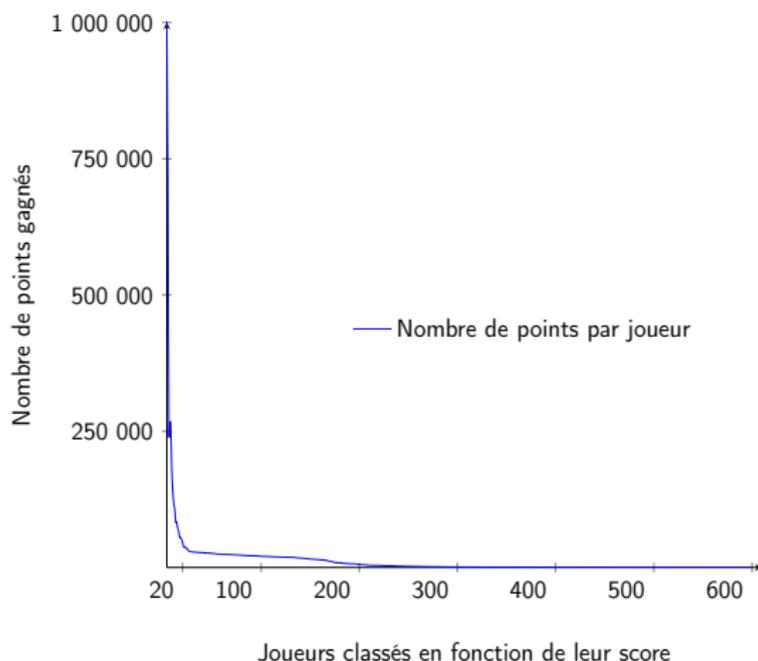


Nombre de joueurs sur *Phrase Detectives* en fonction de leur classement en points
(février 2011 - février 2012)

Theory and Applications of Natural Language Processing, 2013

Using Games to Create Language Resources : Successes and Limitations of the Approach. The People's Web Meets NLP. J. Chamberlain, **K. Fort**, U. Kruschwitz, M. Lafourcade et M. Poesio

Myriadisation : une foule de « non-experts » ? (GWAP (2))



Nombre de joueurs sur *JeuxDeMots* en fonction de leur classement en points
(source : <http://www.jeuxdemots.org/generateRanking-4.php>)

Myriadisation : une foule de « non-experts » ? (MTurk)

Nombre de *Turkers* actifs sur Amazon Mechanical Turk (MTurk) :

- nombre enregistré sur le site : plus de 500 000
- 80 % des tâches (HIT) sont réalisées par les 20 % de *Turkers* les plus actifs [Deneme, 2009]

⇒ réellement actifs : entre 15 059 et 42 912

Computational Linguistics, 37 :2, 2011

Amazon Mechanical Turk : Gold Mine or Coal Mine ?. Last Words editorial. K. Fort, G. Adda et K. B. Cohen

Myriadisation

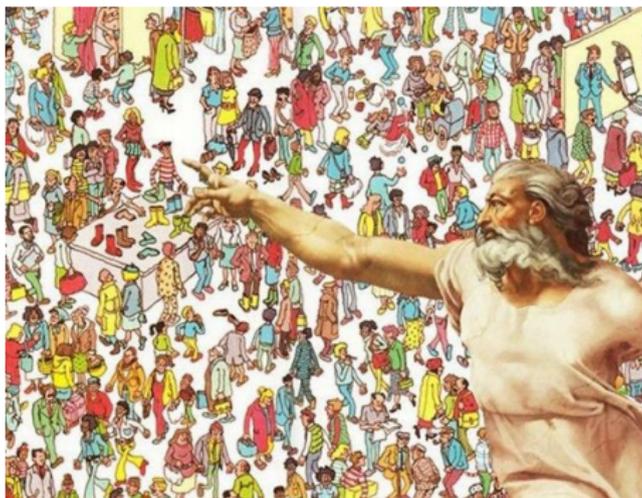
Faire annoter des « non-experts » ?



Myriadisation

Faire annoter des « non-experts » ?

→ Trouver/former des experts (de la tâche) dans la foule



- 1 Introduction
- 2 Myriadiser pour créer des ressources
- 3 Évaluer la qualité de l'annotation manuelle**
 - Accords inter-annotateurs
 - Donner du sens aux résultats
- 4 Gérer la complexité de l'annotation manuelle
- 5 Conclusion

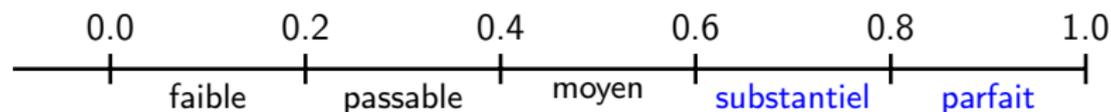
Évaluer l'interprétation humaine ?

On mesure la **fiabilité** de l'annotation
c-à-d si les annotateurs humains prennent des décisions **cohérentes**
en tenant compte du **hasard**

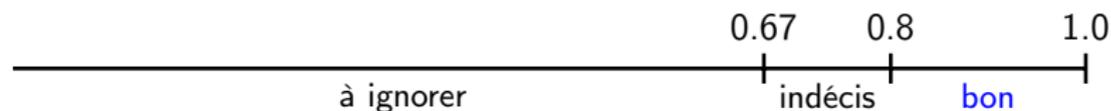
K

Évolution de l'interprétation des κ

[Landis and Koch, 1977]



[Krippendorff, 1980]



[Green, 1997]



Donner du sens aux résultats obtenus

Création d'un outil « Richter » qui :

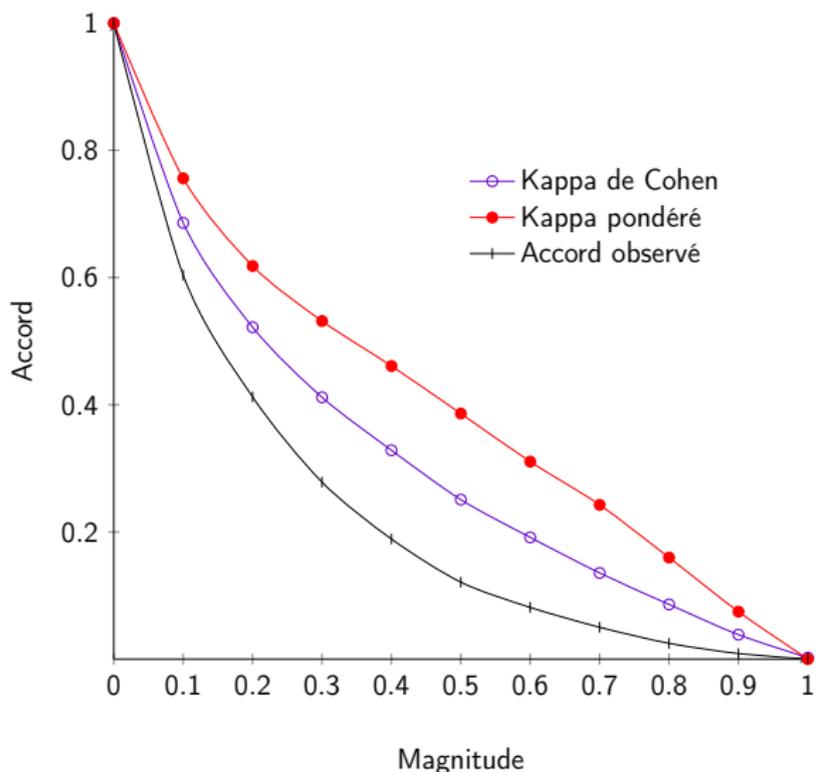
- prend en entrée une annotation de référence (réelle ou générée automatiquement)
- génère des dégradations d'une certaine **magnitude** (de 0 à 1)
- applique une ou des mesures d'accord inter-annotateurs sur chaque ensemble d'annotations (correspondant à une magnitude de dégradation)

COLING 2012 (poster)

Manual Corpus Annotation : Evaluating the Evaluation Metrics. Y. Mathet, A. Widlöcher, **K. Fort**, C. François, O. Galibert, C. Grouin, J. Kahn, S. Rosset et P. Zweigenbaum

Richter sur le corpus TCOF-POS [Benzitoun et al., 2012]

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



1 Introduction

2 Myriadiser pour créer des ressources

3 Évaluer la qualité de l'annotation manuelle

4 Gérer la complexité de l'annotation manuelle

- Exemples d'annotations
- Formaliser les dimensions de complexité
- Le poids du contexte
- Outiller l'annotation manuelle

5 Conclusion

Annoter manuellement

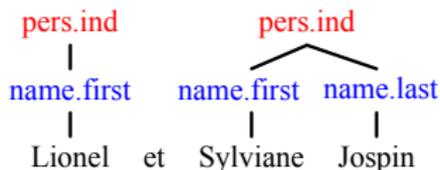
Parties du discours [Marcus et al., 1993] :

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

Renommage de gènes [Fort et al., 2012] :

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.

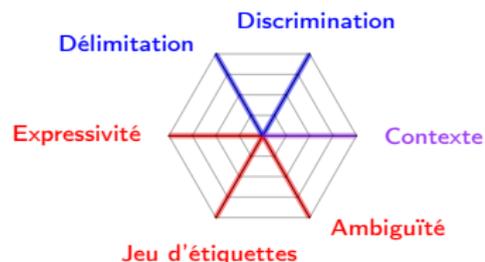
Entités nommées structurées [Grouin et al., 2011] :



⇒ En quoi est-ce facile ou difficile ? Quoi outiller ?

Dimensions de complexité

- 1 Discrimination des unités à annoter
- 2 Délimitation des unités à annoter
- 3 Expressivité du langage d'annotation
- 4 Dimension du jeu d'étiquettes
- 5 Ambiguïté
- 6 Contexte à prendre en compte



- Métriques associées, calculables a priori ou sur un échantillon
- Indépendantes du volume à annoter et du nombre d'annotateurs

COLING 2012

Modeling the Complexity of Manual Annotation Tasks : a Grid of Analysis. K. Fort, A. Nazarenko et S. Rosset

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 [Fort et al., 2012], 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 ./.

$$Discrimination_{PTB_{POS}} = 0$$

Renommage de gènes [Fort et al., 2012] :

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.

$$Discrimination_{Renommage} = 0,95$$

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

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

$$Délimitation_{Renommage} = 0$$

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

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

$$\textit{Expressivité}_{\textit{Renommage}} = 0,25$$

$$\textit{Expressivité}_{\textit{PTB}_{\textit{POS}}} = 0,25$$

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

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>) | 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 | | | | | |

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> | 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.hour.abs</i> (absolute hour), | |
| Amount <i>amount</i> (with unit or general object), including duration | | | <i>time.date.rel</i> (relative date) | <i>time.hour.rel</i> (relative hour) | |

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).

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

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

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

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

$$Ambiguïté_{Res}Renommage = 0,02$$

→ Ne s'applique pas au Penn Treebank (pas de traces).

Degré d'ambiguïté : ambiguïté théorique

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

Définition

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

avec

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

→ Ne s'applique pas aux relations de renommage.

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 !

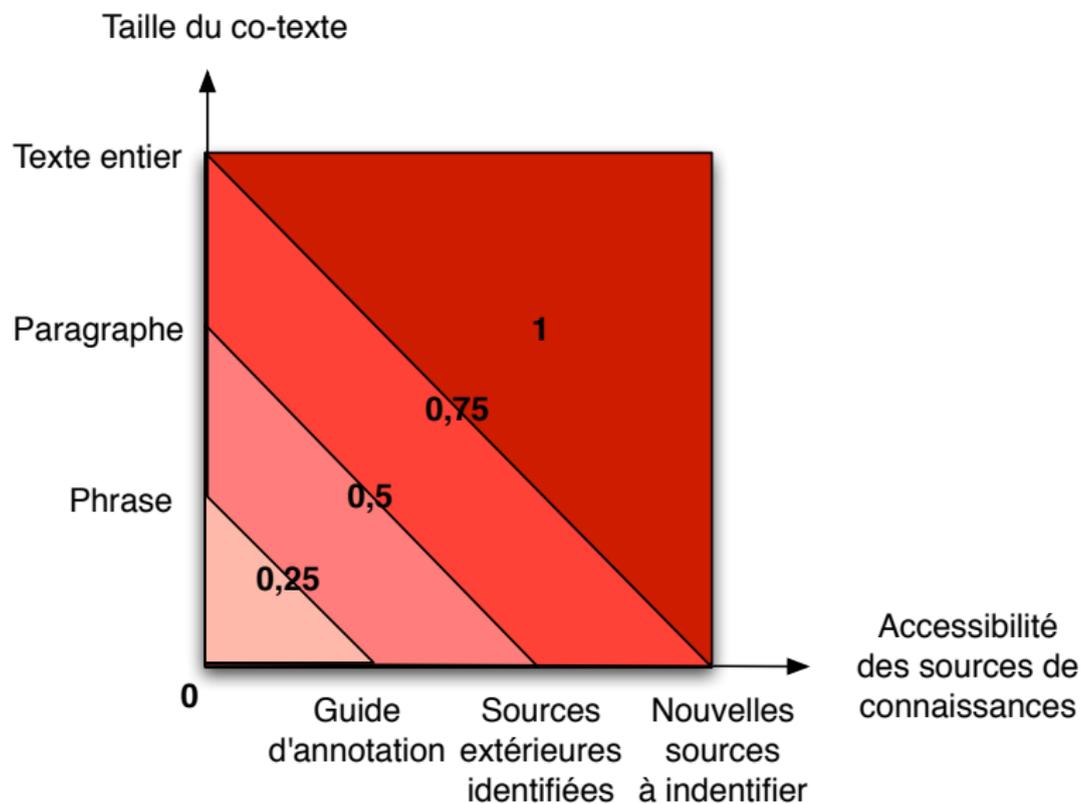
Fabien Lévêque : Et le but du plus breton des **Girondins** . C'est **Joann Gourcuff** qui vient mettre un quatrième but ici au **stade de France** . Le cauchemar continue pour le **VOC** . Quatre à zéro en faveur des **Girondins** .

id=512

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

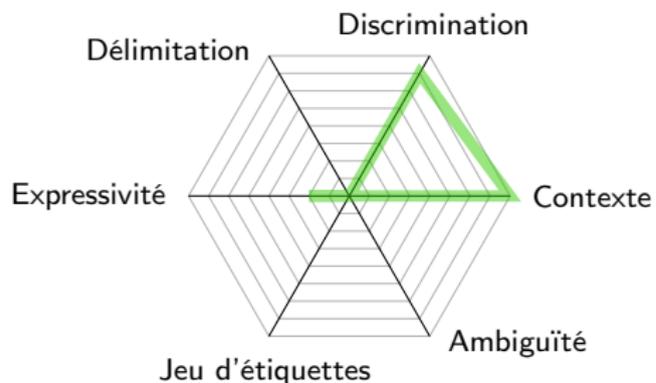
- ▶ guide d'annotation
 - ▶ nomenclatures (Swiss-Prot)
 - ▶ nouvelles sources à trouver (Wikipedia, etc.)

Poids du contexte

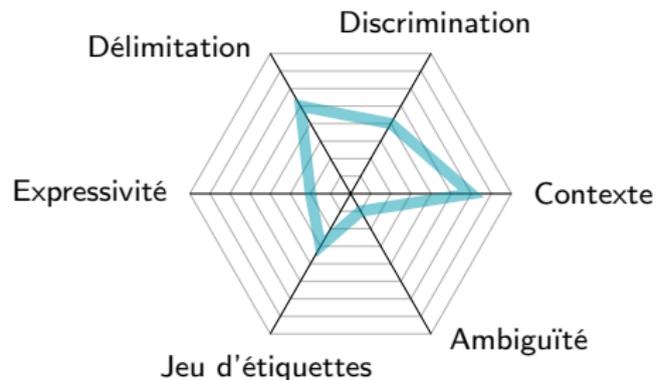


Outiller à bon escient...

Renommage de noms de gènes



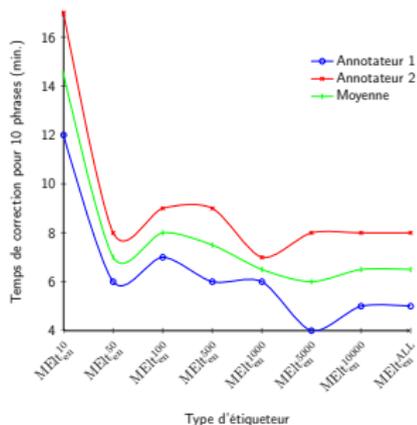
Entités nommées structurées



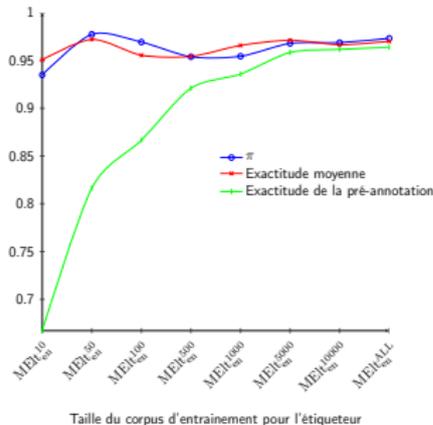
... en fonction du **profil de complexité** de la campagne

Mesurer l'impact de la pré-annotation

- gain en **temps** et en qualité (**accord inter-annotateurs** et **exactitude**)
 - influence de différents **niveaux de qualité** de la pré-annotation
 - **biais** introduit par la pré-annotation
- ... tout en limitant les **effets de la courbe d'apprentissage**



(a) Temps de correction



(b) Qualité de la correction

- 1 Introduction
- 2 Myriadiser pour créer des ressources
- 3 Évaluer la qualité de l'annotation manuelle
- 4 Gérer la complexité de l'annotation manuelle
- 5 **Conclusion**
 - Projet de jeu en cours
 - Vers la myriadisation de tâches linguistiques complexes

Zombilingo : jouer à manger des têtes...

... pour annoter en syntaxe de dépendance

Tâche complexe :

- grand nombre de phénomènes
- instructions contre-intuitives

⇒ gérer la **complexité** :

- en outillant (pré-annotation)
- en formant les joueurs

⇒ **imaginer** des mécanismes :

- pour motiver les joueurs
- les faire revenir
- et créer des annotations de **qualité**



Jouir de la foule est un art [Baudelaire, « Les foules »]

Faire réaliser des tâches complexes nécessite de :

- connaître les dimensions de complexité de la tâche
- pour outiller à bon escient
- former les joueurs et les évaluer
- pour les sélectionner

Obtenir des ressources de qualité nécessite de :

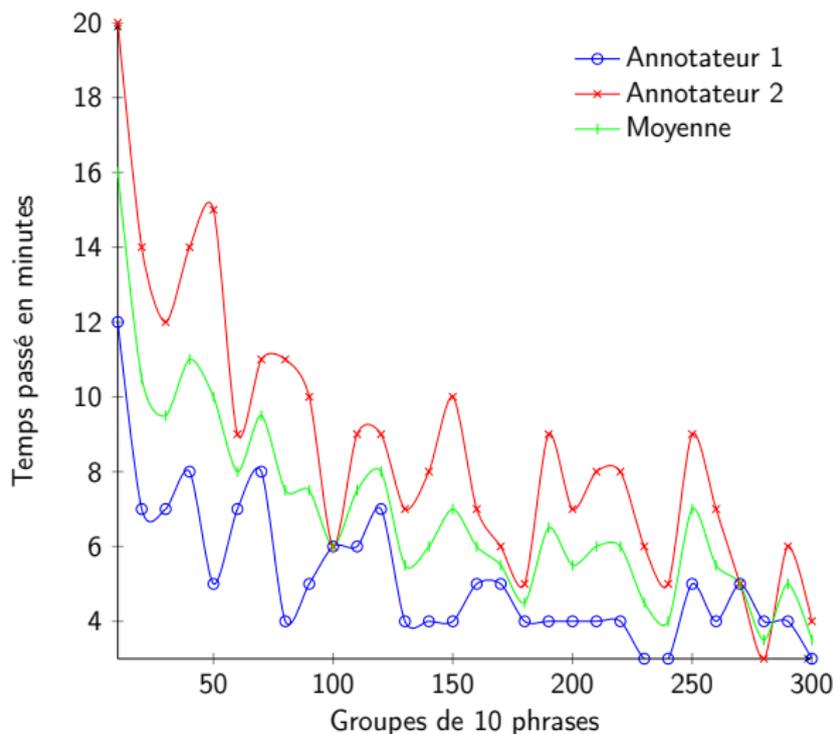
- déterminer les moyens et les formes de l'évaluation
- créer un cercle vertueux



6 Annexes

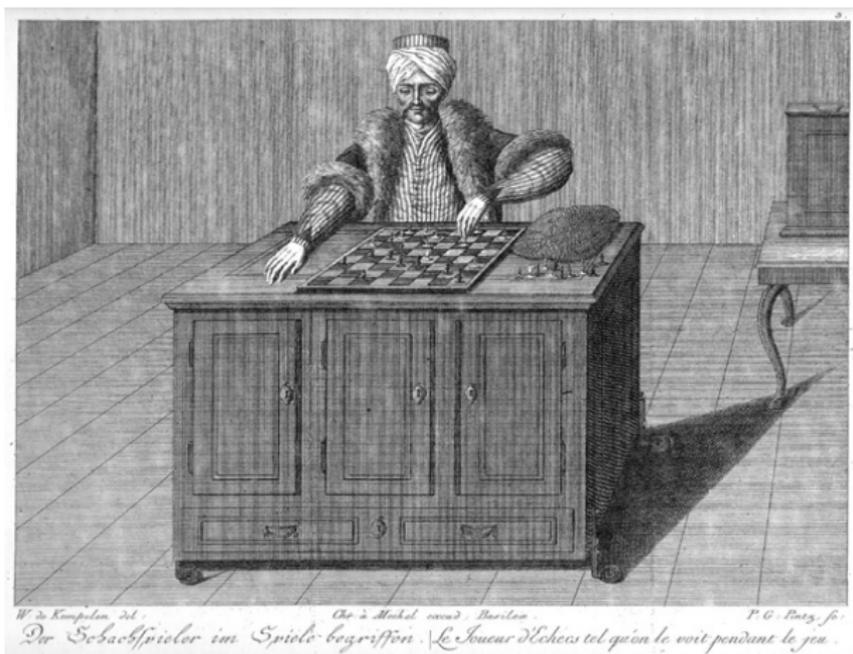
- Pré-annotation
- Amazon Mechanical Turk
- Phrase Detectives
- Zombilingo

Courbe d'apprentissage : annotation morpho-syntaxique du Penn Treebank



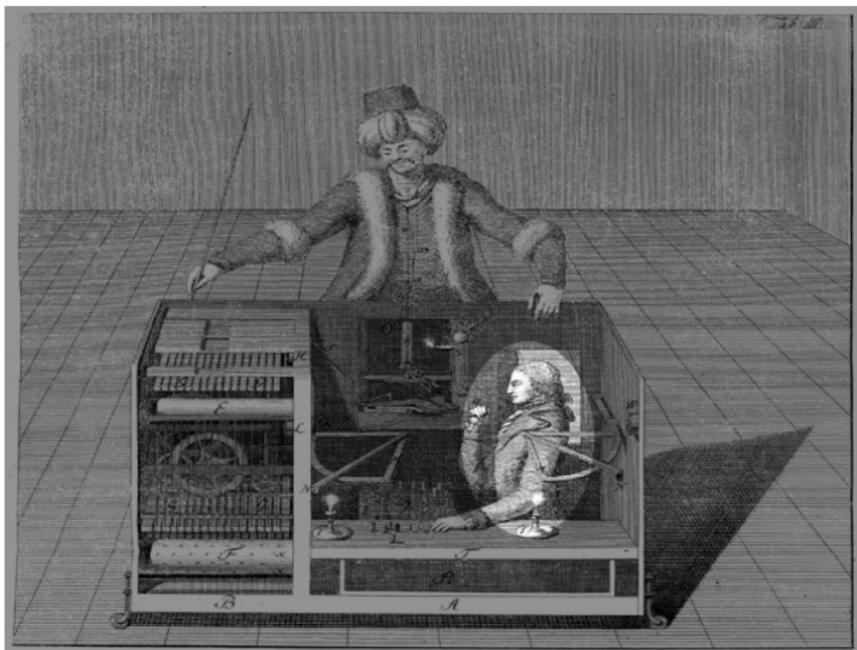
Histoire : le Turc mécanique de von Kempelen

Un joueur d'échecs mécanique créé par J. W. von Kempelen en 1770 :



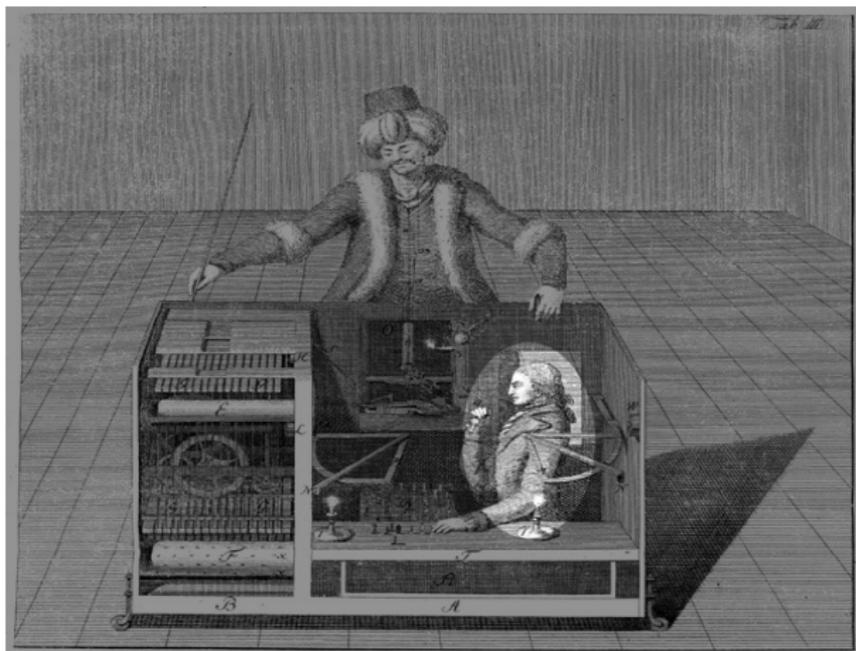
Histoire : le Turc mécanique de von Kempelen

En fait, un maître d'échecs était caché dans la machine :



Histoire : le Turc mécanique de von Kempelen

C'est l'intelligence artificielle **artificielle** !



Amazon Mechanical Turk

MTurk est une plate-forme de **myriadisation** : le travail est *externalisé* via le Web et réalisé par de nombreuses personnes (la *foule*), ici les **Turkers**

amazonmechanicalturk Artificial Intelligence

Your Account | HITs | Qualifications

Introduction | Dashboard | Status | Account Settings

Already have an account? Sign in as a Worker | Requester

Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.
179,373 HITs available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task

Work

Earn money

[Find HITs Now](#)

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a **global, on-demand, 24 x 7 workforce**
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your account

Load your tasks

Get results

[Get Started](#)

Amazon Mechanical Turk

MTurk est une plate-forme de **myriadisation du travail parcellisé** : les tâches sont découpées en sous-tâches (HIT) et leur exécution est payée par les Requesters

amazonmechanicalturk
Artificial Intelligence

Already have an account?
Sign in as a Worker | Requester

Your Account | HITS | Qualifications

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.

179,373 HITS available. [View them now.](#)

Make Money by working on HITs

HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task → **Work** → **Earn money**

Find HITs Now

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your account → **Load your tasks** → **Get results**

Get Started

Amazon Mechanical Turk

MTurk est une plate-forme de **myriadisation du travail parcellisé** : les tâches sont découpées en sous-tâches (HIT) et leur exécution est **payée**.

amazonmechanicalturk
Artificial Intelligence

Already have an account?
Sign in as a Worker | Requester

Your Account | HITS | Qualifications

Introduction | Dashboard | Status | Account Settings

Mechanical Turk is a marketplace for work.

We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.

179,373 HITS available. [View them now.](#)

Make Money
by working on HITS

HITS - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITS now.](#)

As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task → Work → Earn money

Find HITS Now

Get Results
from Mechanical Turk Workers

Ask workers to complete HITS - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Register Now](#)

As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITS completed in minutes
- Pay only when you're satisfied with the results

Fund your account → Load your tasks → Get results

Get Started

Amazon Mechanical Turk

MTurk est une plate-forme de **myriadisation du travail parcellisé** : les tâches sont découpées en sous-tâches (HIT) et leur exécution est **payée**.

The screenshot shows the Amazon Mechanical Turk homepage. At the top, there are navigation tabs for 'Your Account', 'HITs', and 'Qualifications'. A banner states 'Mechanical Turk is a marketplace for work.' Below this, there are two main sections:

Make Money by working on HITs
 HITs - Human Intelligence Tasks - are individual tasks that you work on. [Find HITs now.](#)
As a Mechanical Turk Worker you:

- Can work from home
- Choose your own work hours
- Get paid for doing good work

A flow diagram shows: Find an interesting task (with a list of tasks like 'Find account', 'Load supply chain', 'Analyze online reviews', 'TASKS after government officials', 'Review', 'New ad campaign') → Work (gears icon) → Earn money (dollar sign icon). A 'Find HITs Now' button is at the bottom.

Get Results from Mechanical Turk Workers
 Ask workers to complete HITs - Human Intelligence Tasks - and get results using Mechanical Turk. [Register Now](#)
As a Mechanical Turk Requester you:

- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- **Pay only when you're satisfied with the results**

A flow diagram shows: Fund your account (plus icon) → Load your tasks (gears icon) → Get results (star icon). A 'Get Started' button is at the bottom.

MTurk : un passe-temps pour les Turkers ?

[Ross et al., 2010, Ipeirotis, 2010] montre que :

- Turkers sont avant tout motivés par l'**argent** (91 %) :
 - ▶ 20 % considèrent MTurk comme leur source de revenu primaire ;
 - ▶ 50 % comme leur source de revenu secondaire ;
 - ▶ l'aspect loisir n'est important que pour une minorité (30 %).
- 20 % des Turkers passent plus de 15 h par semaine sur MTurk, et contribuent à 80 % des tâches.
- le salaire horaire moyen observé est **inférieur à 2 \$**.

Phrase Detectives : bon/mauvais joueur

| | System | Good player | Bad player |
|--------------------------|---------------|-------------|--------------|
| ANNOTATIONS | | | |
| Total Annotations: | 1423078 | 4587 | 11018 |
| Average Annotation Time: | 00:00:07 | 00:00:07 | 00:00:04 |
| Total (Ratio) DN: | 955520 (0.67) | 1495 (0.33) | 10935 (0.99) |
| Total (Ratio) DO: | 378256 (0.27) | 2696 (0.59) | 58 (0.01) |
| Total (Ratio) PR: | 79172 (0.06) | 334 (0.07) | 24 (0) |
| Total (Ratio) NR: | 13395 (0.01) | 64 (0.01) | 2 (0) |
| VALIDATIONS | | | |
| Total Validations: | 608982 | 3848 | 5256 |
| Total (Ratio) Agree: | 200174 (0.33) | 1186 (0.31) | 8 (0) |
| Ave Agree Time: | 00:00:09 | 00:00:08 | 00:00:18 |
| Total (Ratio) Disagree: | 408808 (0.67) | 2662 (0.69) | 5248 (1) |
| Ave Disagree Time: | 00:00:08 | 00:00:07 | 00:00:02 |
| OTHER | | | |
| Total Skips: | 51616 | 142 | 26 |
| Skip per annotation: | 0.04 | 0.03 | 0 |
| Total Comments: | 26593 | 229 | 0 |
| Comment per annotation: | 0.02 | 0.05 | 0 |

“The bad player in this case was identified by the speed of annotations and that the only responses were DN in Annotation Mode and Disagree in Validation Mode. The player later confessed to using automated form completion software.” [Chamberlain et al., 2013]

Phrase Detectives : accord inter-annotateurs

| | Expert 1 vs. Expert 2 | Expert 1 vs. Game | Expert 2 vs. Game |
|--------------------|------------------------------|--------------------------|--------------------------|
| Accord général | 94,1 % | 84,5 % | 83,9 % |
| Discourse-new (DN) | 93,9 % | 96,0 % | 93,1 % |
| Discourse-old (DO) | 93,3 % | 72,7 % | 70 % |
| Non-referring (NR) | 100 % | 100 % | 100 % |
| Property (PR) | 100 % | 0 % | 0 % |



ZOMBILINGO



JOUEUR 1
Zombie affamé



340

Invasion - Phrase 3/5



Tu as gagné 334 points et 23 pièces pour le moment

Besoin d'aide pour
cette relation ?
Demande-moi !



ELLE A FAIT MONTER SON ASSISTANT



Trouve le mot associé à la relation SUJET en glissant la main dessus

utiliser
un objet

Récompense pour l'analyse de cette phrase : 40 points et 5 pièces

passer

soumettre

quitter
la partie

-  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.
-  Chamberlain, J., Fort, K., Kruschwitz, U., Lafourcade, M., and Poesio, M. (2013).
Using games to create language resources : Successes and limitations of the approach.
In Gurevych, I. and Kim, J., editors, The People's Web Meets NLP, Theory and Applications of Natural Language Processing, pages 3–44. Springer Berlin Heidelberg.

-  Chamberlain, J., Kruschwitz, U., and Poesio, M. (2009).
Constructing an anaphorically annotated corpus with non-experts :
assessing the quality of collaborative annotations.
In Proceedings of the 2009 Workshop on The People's Web Meets
NLP : Collaboratively Constructed Semantic Resources, People's Web
'09, pages 57–62, Stroudsburg, PA, USA. Association for
Computational Linguistics.
-  Chamberlain, J., Poesio, M., and Kruschwitz, U. (2008).
Phrase Detectives : a web-based collaborative annotation game.
In Proceedings of the International Conference on Semantic Systems
(I-Semantics'08), Graz, Autriche.
-  Cooper, S., Treuille, A., Barbero, J., Leaver-Fay, A., Tuite, K., Khatib,
F., Snyder, A. C., Beenen, M., Salesin, D., Baker, D., and Popović, Z.
(2010).
The challenge of designing scientific discovery games.

In Proceedings of the Fifth International Conference on the Foundations of Digital Games, FDG '10, pages 40–47, New York, NY, USA. ACM.



Deneme (2009).

How many turkers are there?

<http://groups.csail.mit.edu/uid/deneme/>.



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

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.



Gala, N. and Brun, C. (2012).

Propagation de polarités dans des familles de mots : impact de la morphologie dans la construction d'un lexique pour l'analyse d'opinions.

In Actes de Traitement Automatique des Langues Naturelles (TALN 2012), Grenoble.



Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., and Schader, M. (2011).

Managing the crowd : Towards a taxonomy of crowdsourcing processes.

In AMCIS 2011 Proceedings.



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.



Ipeirotis, P. (2010).

The new demographics of mechanical turk.

<http://behind-the-enemy-lines.blogspot.com/2010/03/new-demographics-of-mechanical-turk.html>.



Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popović, Z., et al. (2011).

Crystal structure of a monomeric retroviral protease solved by protein folding game players.

[Nature structural & molecular biology](#), 18(10) :1175–1177.



Krippendorff, K. (1980).

[Content Analysis : An Introduction to Its Methodology](#), chapter 12.

Sage, Beverly Hills, CA., USA.



Lafourcade, M. and Joubert, A. (2008).

JeuxDeMots : un prototype ludique pour l'émergence de relations entre termes.

In Actes de Journées internationales d'Analyse statistique des Données Textuelles (JADT), Lyon, France.



Lafourcade, M., Joubert, A., Schwab, D., and Zock, M. (2011).
Evaluation et consolidation d'un réseau lexical grâce à un assistant ludique pour le mot sur le bout de la langue.

In Actes de Traitement Automatique des Langues Naturelles (TALN), pages 295–306, Montpellier, France.



Landis, J. R. and Koch, G. G. (1977).
The measurement of observer agreement for categorical data.
Biometrics, 33(1) :159–174.



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.



Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and Tomlinson, B. (2010).

Who are the crowdworkers? : shifting demographics in mechanical turk.

In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, CHI EA '10, pages 2863–2872, New York, NY, USA. ACM.