

Annotation pour le Traitement Automatique de la Langue

Ethique, modélisation et pratique

Aurélie Névéol

4 octobre 2022

aurelie.neveol@lisn.upsaclay.fr

Résumé:

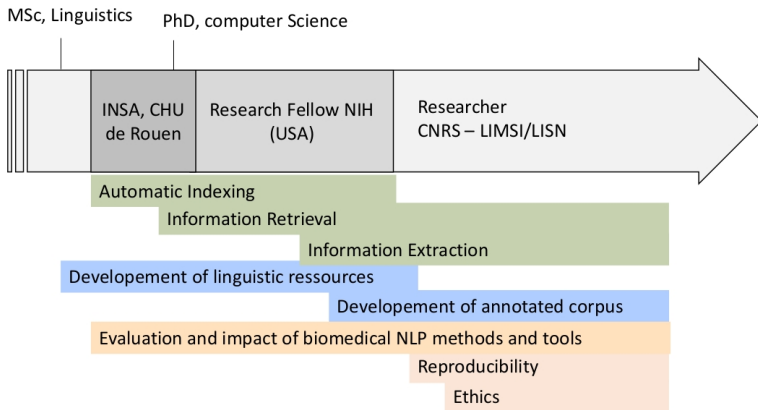
L'annotation est au coeur du TAL

L'annotation est une tâche complexe

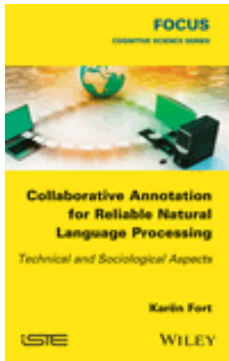
L'annotation a un impact à long terme

Introduction

Qui suis-je?



- La place de l'annotation dans le TAL
- Pourquoi c'est difficile de produire des annotations?
- En pratique comment on fait?



Fort K. 2016. Collaborative Annotation for Reliable Natural Language Processing. ISTE-Wiley

La place de l'annotation dans le TAL

À quoi sert le TAL?

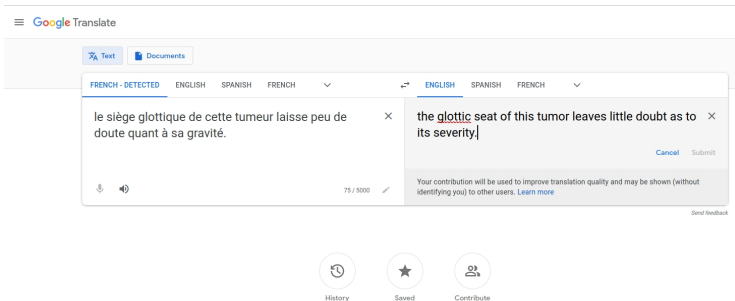
Reconnaissance de la parole



Le TAL, c'est génial !

À quoi sert le TAL?

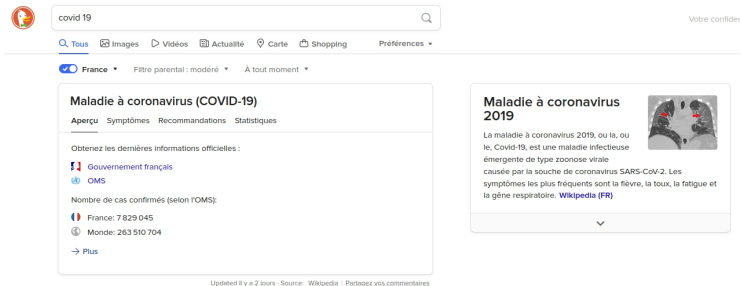
Traduction automatique



The screenshot displays the Google Translate web interface. At the top, the Google Translate logo is visible. Below it, there are tabs for 'Text' and 'Documents'. The main interface is split into two panels. The left panel shows the source text in French: 'le siège glottique de cette tumeur laisse peu de doute quant à sa gravité.' Below the text are icons for voice input and output, and a character count of '75 / 5000'. The right panel shows the translated text in English: 'the glottic seat of this tumor leaves little doubt as to its severity|'. There are 'Cancel' and 'Submit' buttons next to the translation. At the bottom of the right panel, there is a notice: 'Your contribution will be used to improve translation quality and may be shown (without identifying you) to other users. Learn more'. Below the main interface, there are three circular icons: 'History' (a clock), 'Saved' (a star), and 'Contribute' (a person icon).


À quoi sert le TAL?

Extraction d'information



The image shows a search engine interface with the search term "covid 19". The results are displayed in two columns. The left column features a detailed result for "Maladie à coronavirus (COVID-19)", including a summary, official information from the French government and WHO, and confirmed case counts for France (7,829,045) and the world (263,510,704). The right column shows a smaller result for "Maladie à coronavirus 2019" with a chest X-ray image and a brief description of the disease as a zoonotic viral illness caused by SARS-CoV-2.

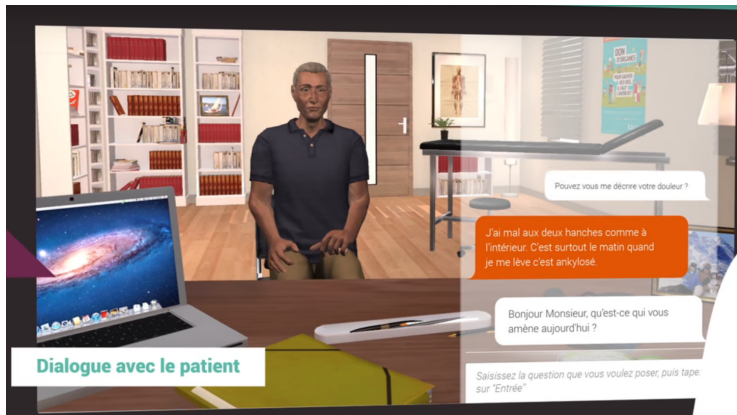
Search results for "covid 19":

- Maladie à coronavirus (COVID-19)**
 - Aperçu Symptômes Recommandations Statistiques
 - Obtenez les dernières Informations officielles :
 - [Gouvernement français](#)
 - [OMS](#)
 - Nombre de cas confirmés (selon l'OMS):
 - France: 7 829 045
 - Monde: 263 510 704
 - [→ Plus](#)
- Maladie à coronavirus 2019**
 - 
 - La maladie à coronavirus 2019, ou la, ou le, Covid-19, est une maladie infectieuse émergente de type zoonose virale causée par la souche de coronavirus SARS-CoV-2. Les symptômes les plus fréquents sont la fièvre, la toux, la fatigue et la gêne respiratoire. [Wikipedia \(FR\)](#)

Updated il y a 2 jours - Source: [Wikipedia](#) | [Partagez vos commentaires](#)

À quoi sert le TAL?

Agents conversationnels



À quoi sert le TAL?

Etc.

Comment ça marche?

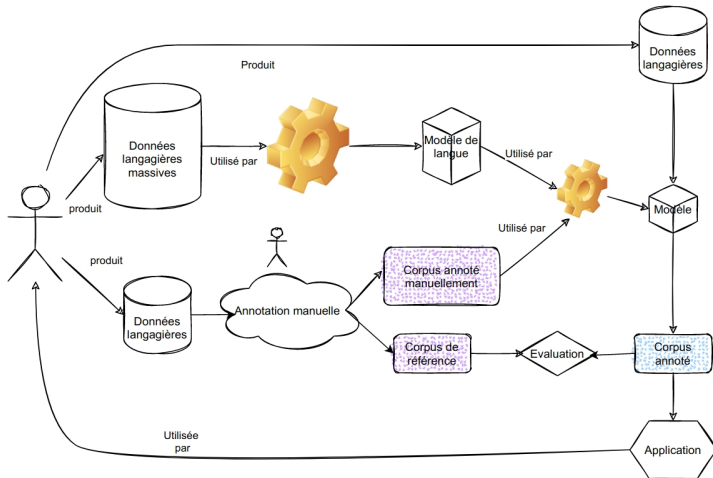


Image: K. Fort

Comment ça marche?

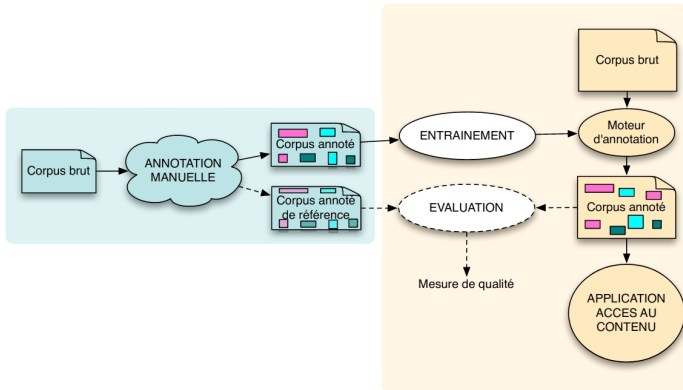
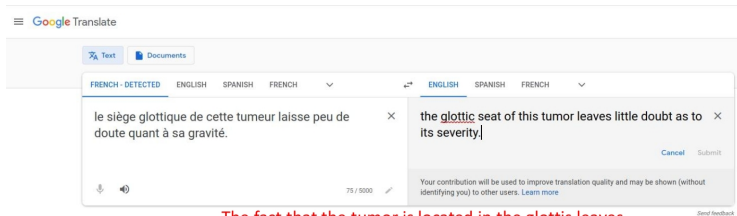


Image: K. Fort

Comment sont créées les annotations?

Par des citoyens/ utilisateurs actifs



The fact that the tumor is located in the glottis leaves little doubt as to its severity.



History



Saved



Contribute

Comment sont créées les annotations?

Par des citoyens/ utilisateurs a posteriori

Case Reports > [Rev Med Liege](#). 2020 Dec;75(12):767-769.

[Cervical spondylolysis : a rare fortuitous finding]

[Article in French]

[C Bodet](#)¹, [P J Bruyère](#)², [A Ferré](#)³, [J F Kaux](#)¹, [A C Courtois](#)⁴

Affiliations + expand

PMID: 33331698

Abstract in English, [French](#)

A young patient consulted in our physical and rehabilitation medicine department following the onset of pain on the scapula area and at the base of his right upper limb after carrying a heavy load. After a couple of weeks, the patient also developed cervical pain. Fortuitously, the cervical scanner displayed a right C6 spondylolysis. Further evaluation by bone scan confirmed that this lesion was not recent and so probably not the cause of the symptoms. Cervical isthmic spondylolysis is a rare condition, much more common at the lumbar level and often ignored at the cervical one. The etiology, pathophysiology, imaging and treatment options for this cervical pathology are discussed in this article.

Comment sont créées les annotations?

Par le jeu (sciences participatives)

- Co-référence

<https://anawiki.essex.ac.uk/phrasedetectives/>

PHRASE • **DETECTIVES**

- Syntaxe <https://zombiludik.org/>

ZOMBILUDIK

- Expressions polylexicales <http://rigor-mortis.org/>

RIGORMORTIS

Comment sont créées les annotations?

Par des annotateurs "experts"

- Définition de l'expertise ?
- La formation et les conditions de travail des annotateurs ont un impact scientifique et éthique [Fort et al. 2011]

Fort K, Adda G, Cohen KB. Amazon Mechanical Turk: Gold Mine or Coal Mine? Computational Linguistics, 37(2):413-420 (2011)

Les corpus annotés continuent d'être utilisés des **décennies** après leur création

- Penn Tree Bank (Markus et al. 1993)
- French Tree Bank (Abeillé, et al. 2003)
- ...

crédit transparent: K. Fort

Pourquoi c'est difficile?

Transcription du compte-rendu de consultation de Mme Annick Dupont par le Dr. Faitsemblant:

1. Écouter l'enregistrement

`https://perso.limsi.fr/neveol/CRpourTranscription.m4a`

2. Produire un texte écrit correspondant

On cherche à identifier le genre des utilisateurs d'un forum d'après leurs écrits. Proposez un protocole d'extraction d'information.

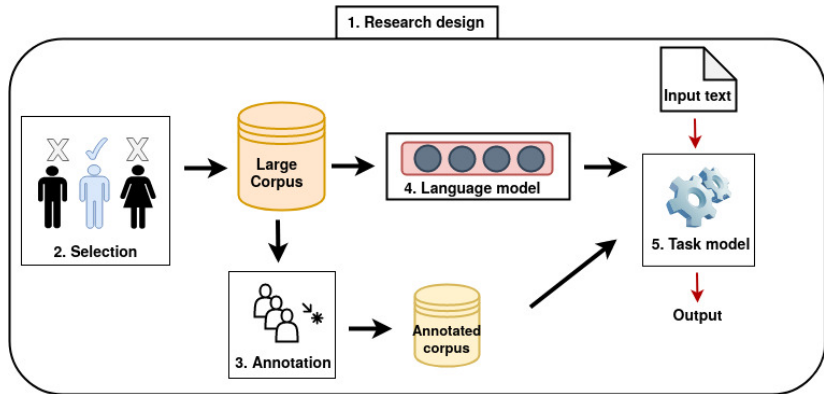
Annoter les expressions temporelles, les maladies, les procédures, les personnes, les médicaments:

J'ai revu en consultation le 24 Mai 2022 Mme Dupont qui est en attente d'une transplantation hépatique. Elle a pris 8 kilos en un mois avec la présence d'une ascite tendue aujourd'hui et des oedèmes des membres inférieurs essentiellement vespéraux. Une ponction d'ascite est programmée à l'HEGP. La patiente bénéficie d'une supplémentation injectable par VENOFER. Le bilan électrolytique est tout à fait satisfaisant compte tenu des fortes doses de diurétiques reçues et malgré une valeur de potassium proche de la limite haute, à 4,37 mmol/l.

En pratique, comment on fait?

1. Préparer la campagne d'annotation
 - Envisager les aspects éthiques
 - Planifier les aspects techniques et théoriques
2. Conduite de la campagne d'annotation
 - Un processus itératif
 - Suivi technique et scientifique
3. Documentation

Attention aux sources de biais inhérentes à l'annotation...



Hovy D, Prabhumoye S. (2021). Five sources of bias in natural language processing. *Language and Linguistics Compass*, e12432. <https://doi.org/10.1111/lnc3.12432>

Le problème est-il pertinent et motivé correctement?

- Qui contribue aux décisions de modélisation?
 - L'équipe est-elle inclusive, en termes de parties prenantes, de profil des personnes?
- Quel est l'équilibre des forces?
 - Ingénieurs, financeurs, utilisateurs
 - Quelle gestion des conflits d'intérêts?
- Quelles sont les contraintes techniques?
 - Contenu et nature des données (surexposition?)
 - Disponibilité des données (sur-généralisation)
- ...

- Quand et comment contacter un comité d'éthique?
- La langue implique des personnes humaines
 - Sources → notion de confidentialité, atteinte potentielle à la représentation des personnes
 - Traitement de corpus → compensation des annotateurs
 - Impact sur les utilisateurs → direct et indirect

Noble SU. Algorithms of oppression: how search engines reinforce racism. NYU Press (2018)

Larson B Gender as a Variable in Natural-Language Processing: Ethical Considerations. Proc. ACL Workshop on Ethics in Natural Language Processing 1-11. (2017).

Impact des données sur l'évaluation?

- Similarité entre les corpus d'entraînement et de test
 - Notions d'adaptation en domaine, de surapprentissage
 - Impact potentiel de la redondance
- Evaluation de la capacité des outils à mémoriser et à généraliser
 - Par rapport à une situation réaliste d'utilisation?

Elangovan A, He J, Verspoor K. Memorization vs. Generalization : Quantifying Data Leakage in NLP Performance Evaluation. Proc. EACL 2021: 1325-1335

Processus d'annotation collaborative

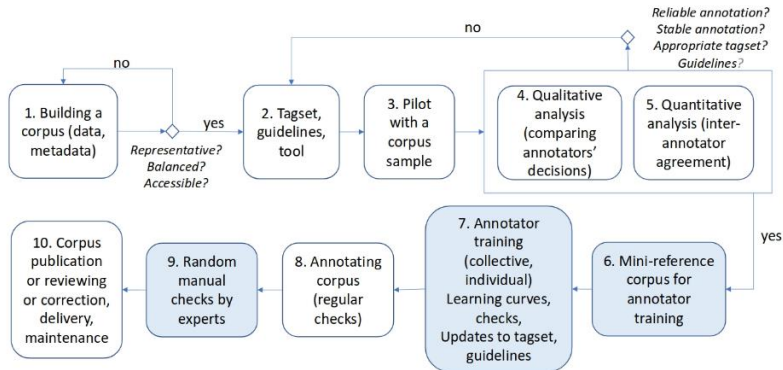


schéma: E. Volodina

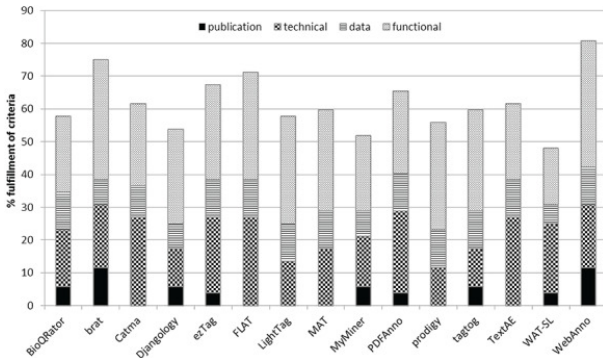
- Quel contenu?
 - Granularité (phrase, document, dossier)
 - Volume (notion de représentativité et de temps d'annotation)
- Modes de sélection
 - Echantillonnage [Campillos et al. 2017]
 - Signaux faibles [Névéol et al. 2011]
 - Maitriser la redondance [Rabary et al. 2015]

Une description précise des annotations à réaliser

- Définitions des notions à annoter
- Illustration en corpus par des exemples
- Présentation de cas limites

Choix d'un outil d'annotation

Evaluation of **installable** (2 hrs), **workable** (user defined schema), tools



Neves M, Ševa J. An extensive review of tools for manual annotation of documents. Brief Bioinform. 2021 Jan 18;22(1):146-163.

Quelques critères importants:

- Possibilité d'utiliser une pré-annotation?
- Gestion d'une campagne avec de multiples annotateurs?

Tout au long de la campagne

- Produire des annotations soigneuses, complètes
 - Importance des outils, de la configuration
- Faire intervenir de multiples annotateurs
 - Formation
 - Recouvrement du travail d'annotation
 - Rôles
- Arriver à un consensus
 - Éventuelles mises à jour du guide

Comment évaluer la qualité des annotations?

- Objectif: déterminer si les annotations réalisées sont *correctes*
- Challenge: pas de "vérité terrain"
 - les catégories annotées s'appuient souvent sur le jugement humain
- Au total: on ne peut pas mesurer la correction directement.
- Solution: mesurer la *consistance* de l'annotation (l'accord entre annotateurs)
 - Hypothèse: plus les annotateurs sont d'accord, plus la qualité des annotations est élevée

Rationale for measuring agreement

Agreement among annotators

↓ *demonstrates*

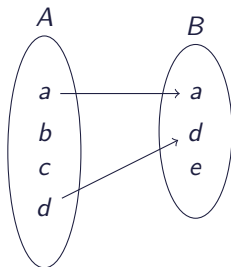
Reliable annotation process

↓ *necessary but not sufficient for*

Correct annotations

Illustration: Artstein 2009

Comment calculer un accord inter-annotateur?



- Annotations en commun
 - Prise en compte de l'aléatoire (erreur, oubli, ajout)
 - Prise en compte de biais et de distributions

Ron Artstein and Massimo Poesio. 2008. Survey Article: Inter-Coder Agreement for Computational Linguistics. Computational Linguistics, 34(4):555–596.

Accord entre paires d'annotateurs:

- Précision, rappel, F-mesure
- Cohen's κ (Cohen, 1960)
- Fleiss's κ (Fleiss, 1971)

Accord multi-annotateurs:

- α (Krippendorff, 1980)

Ron Artstein and Massimo Poesio. 2008. Survey Article: Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4):555–596.

Interpréter l'accord inter-annotateur

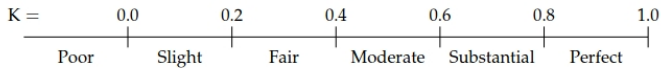


Figure 1
Kappa values and strength of agreement according to Landis and Koch (1977).

Illustration: Artstein & Poesio 2008

Toutes les tâches d'annotations ne sont pas équivalentes

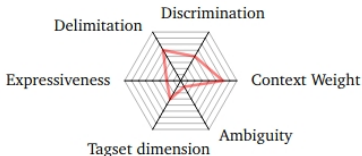
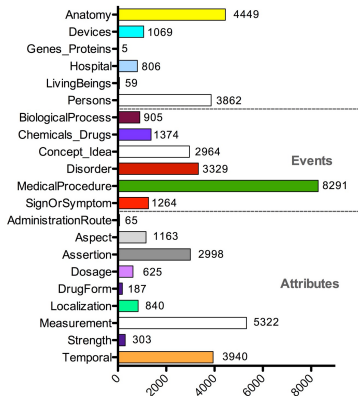


Illustration: Fort et al. 2012

Karën Fort, Adeline Nazarenko, and Sophie Rosset. 2012. Modeling the Complexity of Manual Annotation Tasks: a Grid of Analysis. Proc COLING 2012:895–910.

Illustration: corpus MERLOT, entités

Distribution



Accord inter-annotateur (F1)

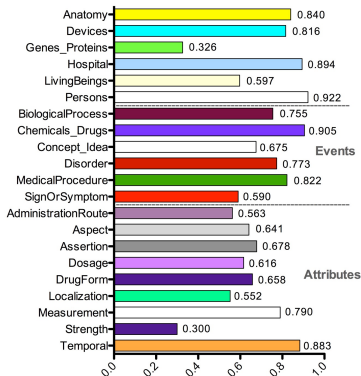
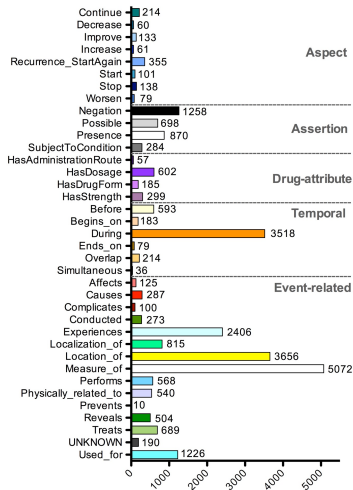
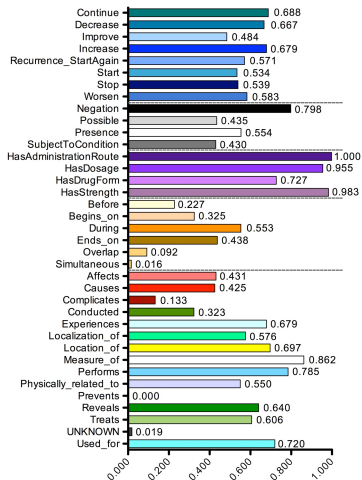


Illustration: corpus MERLOT, relations

Distribution



Accord inter-annotateur (F1)



Documenter le corpus annoté

- Provenance et disponibilité
- Condition d'utilisation: confidentialité, droits
 - Certains type d'information sont considérés comme sensibles (santé, religion, ...)
- Description détaillée
 - Langue, volume
 - Méthode de sélection et de collecte
 - Mesure de qualité, y compris biais potentiels

Bender E and Friedman B Data Statements for NLP: Toward Mitigating System Bias and Enabling Better Science. Transactions of the Association for Computational Linguistics 6:587-604. (2018).

À retenir:

L'annotation est au coeur du TAL

L'annotation est une tâche complexe

L'annotation a un impact à long terme

