

Ce qu'Amazon Mechanical Turk fait
à la recherche :
l'exemple du Traitement Automatique des Langues

Karën Fort

7 mars 2016



D'où je parle



Ce que je ne suis pas :

- ▶ sociologue
- ▶ linguiste
- ▶ spécialiste de l'apprentissage (*machine learning*)



je suis spécialiste en création de ressources langagières pour le traitement automatique des langues (TAL)

- ▶ construire de meilleures ressources, en outillant plus efficacement
- ▶ mieux comprendre les processus à l'œuvre pour améliorer les systèmes de TAL :
 - ▶ apprenabilité par l'humain vs apprenabilité par les systèmes

D'où je parle



Ce que je ne suis pas :

- ▶ sociologue
- ▶ linguiste
- ▶ spécialiste de l'apprentissage (*machine learning*)



je suis spécialiste en création de ressources langagières pour le traitement automatique des langues (TAL)

- ▶ construire **de manière éthique** de meilleures ressources, en outillant plus efficacement
- ▶ mieux comprendre les processus à l'œuvre pour améliorer les systèmes de TAL :
 - ▶ apprenabilité par l'humain vs apprenabilité par les systèmes

Traitement Automatique des Langues (TAL) : l'ogre des données

Des applications dans nos vies

Dans la boîte noire

Chères ressources langagières

La myriadisation comme « solution »

Amazon Mechanical Turk : la solution ou le problème ?

TATA (« There Are Thousands of Alternatives »)

Le TAL, une technoscience omniprésente...

«Traduction» automatique

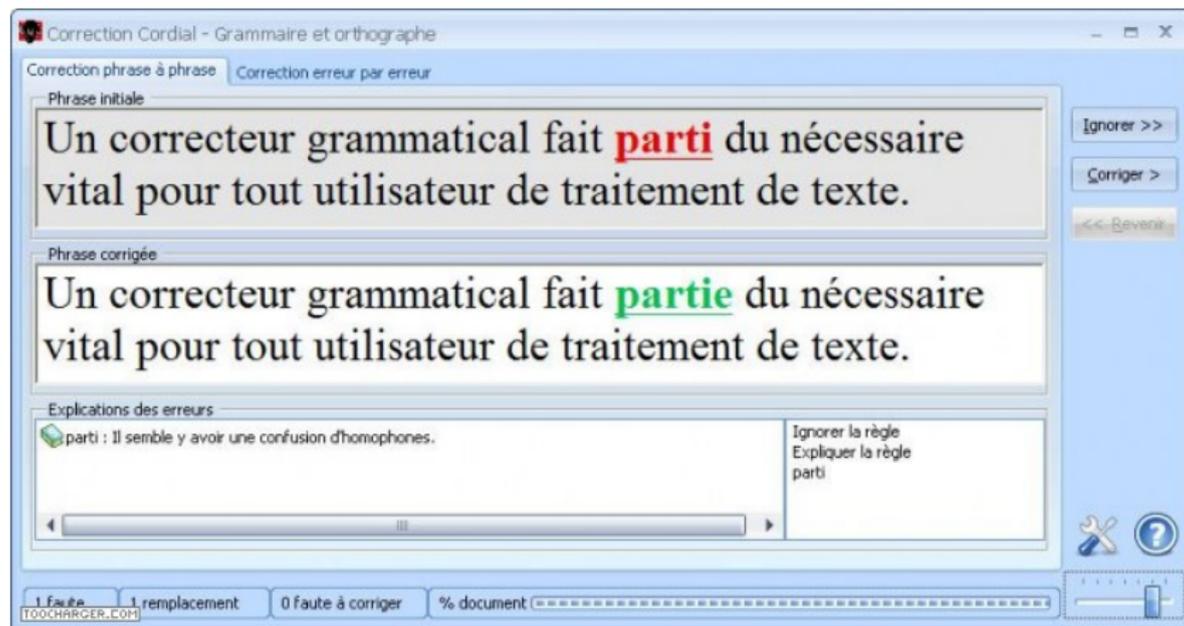


The screenshot shows the Google Translate web interface. At the top left is the Google logo, and at the top right is a 'Connexion' button. Below the logo is the word 'Traduction' in red. On the right side of the interface, there is a red 'G+' icon and a small square icon with a plus sign. The main area has a language selection bar with 'Anglais', 'Français', 'Islandais', and 'Détection la langue' (with a dropdown arrow). Below this is a bidirectional arrow icon. The source language is set to 'Français' and the target language is 'Anglais', with a dropdown arrow for 'Arabe'. A blue 'Traduire' button is on the right. The input text on the left is: "Un algorithme, c'est tout simplement une façon de décrire dans ses moindres détails comment procéder pour faire quelque chose. Il se trouve que beaucoup d'actions mécaniques, toutes probablement, se prêtent bien à une telle décorcation. Le but est d'évacuer la pensée du calcul, afin de le rendre exécutable par une machine numérique (ordinateur, ..). On ne travaille donc qu'avec un reflet numérique du système réel avec qui l'algorithme interagit." (Gérard Berry). The output text on the right is: "An algorithm is simply a way of describing in great detail how to do something. It is a lot of mechanical action, probably all, lend themselves well to such decoration. The goal is to evacuate the thought of the calculation to make it executable by a digital machine (computer, ..). so we only work with a digital reflection of the real system with which the algorithm interacts." (Gérard Berry). At the bottom right of the output area, there is a 'Suggérer une modification' link with a pencil icon.

<https://translate.google.com/>

Le TAL, une technoscience omniprésente...

Correction orthographique et grammaticale



The screenshot shows a software window titled "Correction Cordial - Grammaire et orthographe". It features two tabs: "Correction phrase à phrase" (selected) and "Correction erreur par erreur".

Phrase initiale:
Un correcteur grammatical fait **parti** du nécessaire vital pour tout utilisateur de traitement de texte.

Phrase corrigée:
Un correcteur grammatical fait **partie** du nécessaire vital pour tout utilisateur de traitement de texte.

Explications des erreurs:
parti : Il semble y avoir une confusion d'homophones.

On the right side, there are buttons: "Ignorer >>", "Corriger >", and "<< Revenir". At the bottom right, there are icons for a wrench and a question mark.

Footer:
1 faute, 1 remplacement, 0 faute à corriger, % document. A progress bar is visible below the footer.

<http://www.synapse-fr.com/>

Le TAL, une technoscience omniprésente...

Extraction d'entités nommées

<ENAMEX TYPE="PERSON">Henri</ENAMEX> a acheté **<NUMEX TYPE="QUANTITY">300</NUMEX>** actions de la société **<ENAMEX TYPE="ORGANIZATION">AMD</ENAMEX>** en **<TIMEX TYPE="DATE">2006</TIMEX>**.

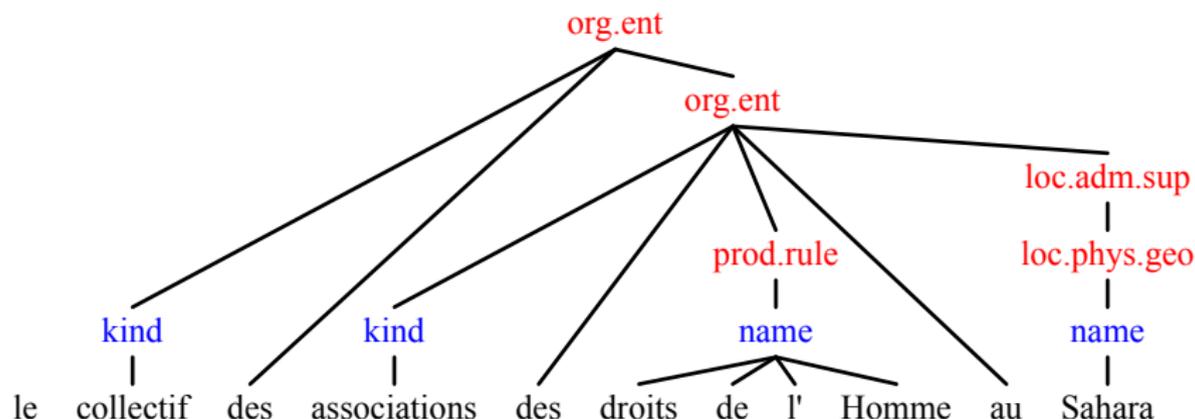
http://fr.wikipedia.org/wiki/Reconnaissance_d%27entit%C3%A9s_nomm%C3%A9es

Le TAL, une technoscience omniprésente...

Extraction d'entités nommées

<ENAMEX TYPE="PERSON">Henri</ENAMEX> a acheté **<NUMEX TYPE="QUANTITY">300</NUMEX>** actions de la société **<ENAMEX TYPE="ORGANIZATION">AMD</ENAMEX>** en **<TIMEX TYPE="DATE">2006</TIMEX>**.

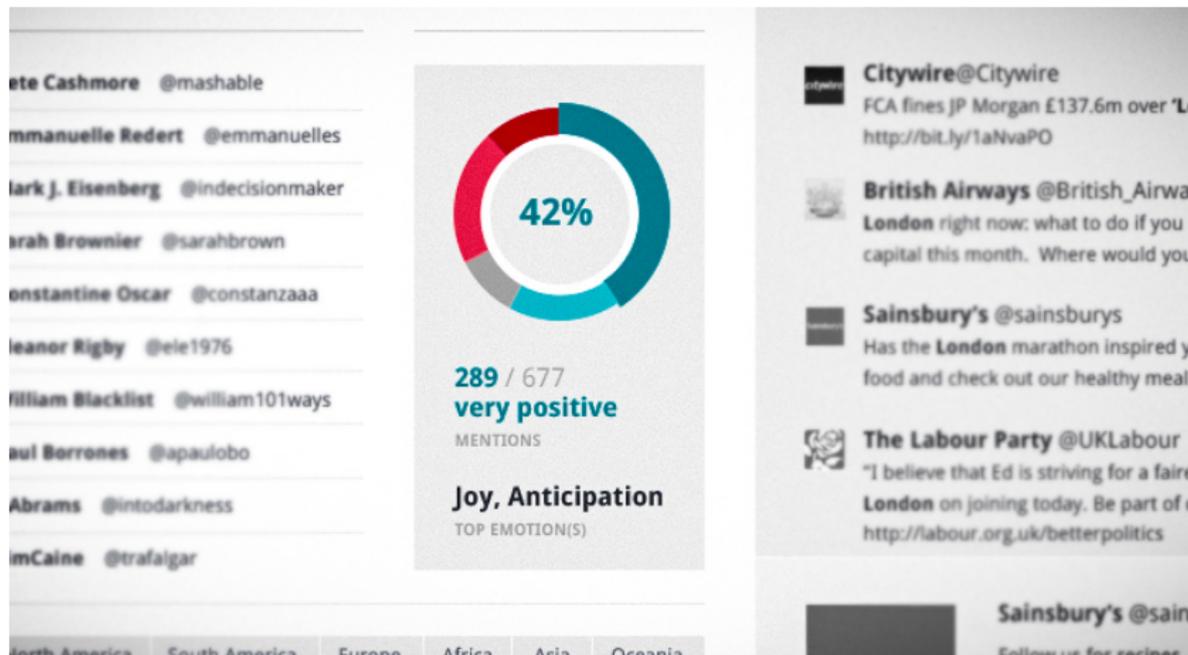
http://fr.wikipedia.org/wiki/Reconnaissance_d%27entit%C3%A9s_nomm%C3%A9es



[Grouin et al., 2011]

Le TAL, une technoscience omniprésente...

Analyse de sentiments



<http://www.sentoapp.com/semantic-analysis.php>

Les mots sont importants : algorithmes et conséquences

**Aucun de nos clics, de nos tweets
ou de nos « like » n'y échappe : les
algorithmes recoupent toutes nos
données pour mieux nous cerner.**

[D. Cardon, Télérama]

<http://www.telerama.fr/idees/les-algorithmes-sont-ils-vraiment-tout-puissants,133304.php>

→ algorithme = tracker (récupération de données) + recoupement

Les mots sont importants : algorithmes et conséquences

Un algorithme de Google bat un humain au go : une question de confiance

Par Camille Gévaudan — 28 janvier 2016 à 17:52

<http://www.liberation.fr/futurs/2016/01/28/>

[un-algorithme-de-google-bat-un-humain-au-go-une-question-de-confiance_1429599](#)

→ algorithme = programme (créé à partir de différents algorithmes)

Les mots sont importants : algorithmes et conséquences



Courrier International

A la une de **Courrier international** cette semaine : ces algorithmes qui nous gouvernent.

En kiosque à partir du 24 septembre <http://bit.ly/1Ffa8wU>

September 23, 2015

<https://www.facebook.com/courrierinternational.com/videos/10153305876269888/>

→ algorithme = pouvoir ?

Les mots sont importants : algorithmes et conséquences



Courrier International

A la une de **Courrier International** cette semaine : ces algorithmes qui nous gouvernent.

En kiosque à partir du 24 septembre <http://bit.ly/1Ffa8wU>

September 23, 2015

<https://www.facebook.com/courrierinternational.com/videos/10153305876269888/>

→ algorithme = pouvoir ?

« Boîte noire » : quelque chose qui n'est plus ouvert, donc plus remis en question [Latour, 1987]

Les mots sont importants : algorithmes et conséquences



Un algorithme n'est pas doué de raison. Ce n'est pas une entité magique. Il n'est ni sournois, ni bienveillant. Simplement composé d'une suite d'instructions élémentaires qui permettent de résoudre un problème. Ceci lui permet d'être exécuté par un ordinateur, capable d'effectuer des opérations infiniment rapidement, de stocker et manipuler des volumes pharaoniques de données, et qui, utilisé à bon escient, est une incroyable annexe aux capacités de calcul de nos cerveaux.

[Anne-Marie Kermarrec (Inria), Libération]

<http://www.liberation.fr/evenements-libe/2014/03/20/>

[les-algorithmes-nous-envahissent-apprenons-a-les-connaître_988330](#)

Les mots sont importants : algorithmes et conséquences

Les algorithmes ne sont pas une fatalité, juste des outils. Cependant, si nous en sommes responsables collectivement, il semble que nous soyons analphabètes individuellement, et pour cause : qui connaît les algorithmes à part les informaticiens ? Il est de notre devoir d'acquérir un minimum de culture à leur sujet, précisément pour ne pas se fourvoyer sur leur puissance et les utiliser efficacement et en connaissance de cause.

[Anne-Marie Kermarrec (Inria), Libération]

<http://www.liberation.fr/evenements-libe/2014/03/20/>

les-algorithmes-nous-envahissent-apprenons-a-les-connaître_988330

« Le TAL, ça marche pas » : l'analyse de CV

« « ben pourquoi il ressort pas le nom ? On le voit bien, là, c'est écrit en gras ! Il marche pas, votre système ». Le dialogue qui s'ensuit est souvent difficile. Je commence par expliquer que le gars du CV s'appelant Marin Martin, il a un prénom qui pourrait être un nom de fonction et un patronyme qui pourrait être un prénom, alors c'est compliqué pour une machine, que le gras est souvent signe de... n'importe quoi dans un CV [...]»

[H. de Mazancourt, Eptica]

<http://www.ethique-et-tal.org/2015/11/09/le-tal-ca-marche-pas/>

La langue naturelle : ambiguïtés à tous les étages

- ▶ transcription :
 - ▶ mon beau-frère **et** ma **sœur** vs mon beau-frère **est** **masseur** vs mon beau-frère **hait** ma **sœur**
- ▶ découpage en « mots » :
 - ▶ l' arbre vs aujourd' hui
- ▶ morpho-syntaxe (nom vs verbe vs adjectif) :
 - ▶ la **belle ferme** la **porte**
- ▶ syntaxe :
 - ▶ je vois la dame **avec** des lunettes
- ▶ résolution (sémantique) d'anaphores :
 - ▶ Le bus a renversé un passant, je **l** 'ai entendu freiner
 - ▶ Le bus a renversé un passant, je **l** 'ai entendu crier

Au cœur du TAL : les ressources langagières

- ▶ systèmes à base de règles
 - ▶ définies par l'humain (linguistes)
 - ▶ entrées manuellement

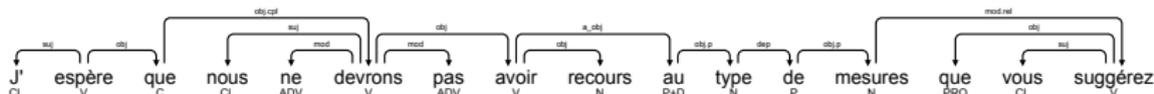
→ **grammaires, lexiques**

```
mirador      nc      [pred='mirador____1<Suj:(sn),Objde:(de-sn|de-sinf),Objà:(à-sinf)>',cat=nc,@ms] mirador____1      Default ms
miradors     nc      [pred='mirador____1<Suj:(sn),Objde:(de-sn|de-sinf),Objà:(à-sinf)>',cat=nc,@mp] mirador____1      Default mp
mirage       nc      [pred='mirage____1<Suj:(sn),Objde:(de-sn|de-sinf),Objà:(à-sinf)>',cat=nc,@ms] mirage____1      Default ms
mirages      nc      [pred='mirage____1<Suj:(sn),Objde:(de-sn|de-sinf),Objà:(à-sinf)>',cat=nc,@mp] mirage____1      Default mp
```

[Lefff, [Sagot, 2010]]

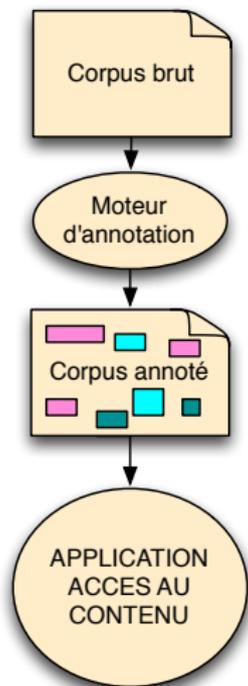
- ▶ systèmes basés sur les données (~ 90 % aujourd'hui)
 - ▶ apprentissage (en général) supervisé (par ex. *deep learning*)
 - ▶ à partir d'exemples (rédigés et/ou annotés par des humains)
 - ▶ algorithmes statistiques (pensés par des humains)

→ **corpus annotés** (et lexiques)

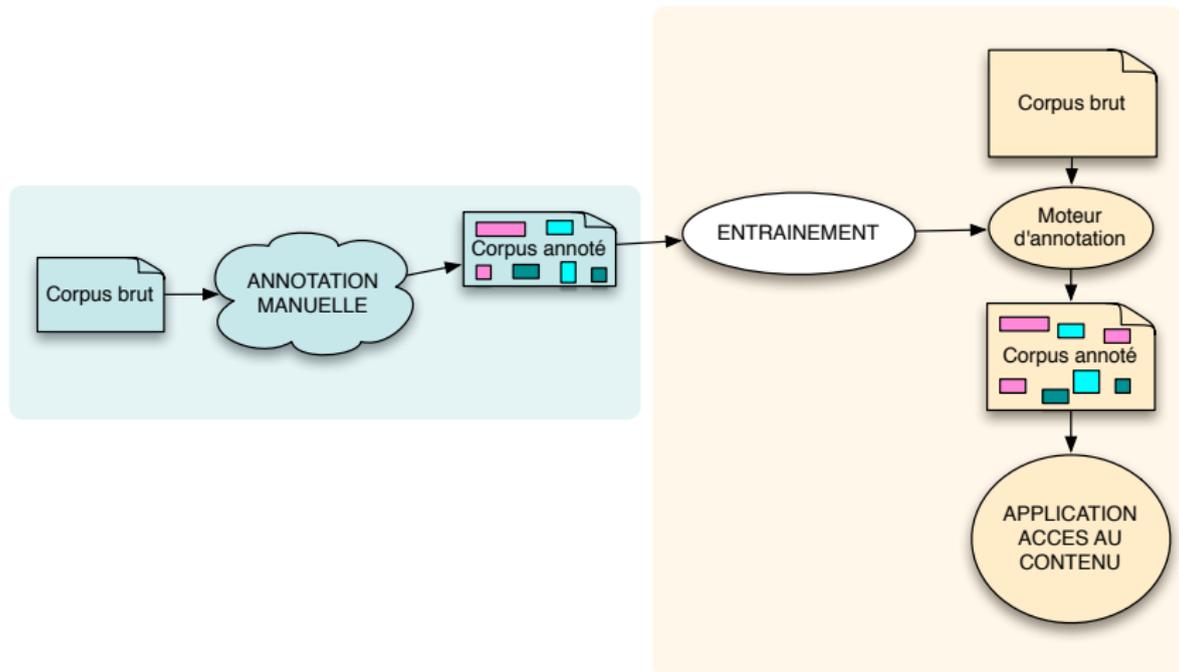


[Annotation en syntaxe de dépendances]

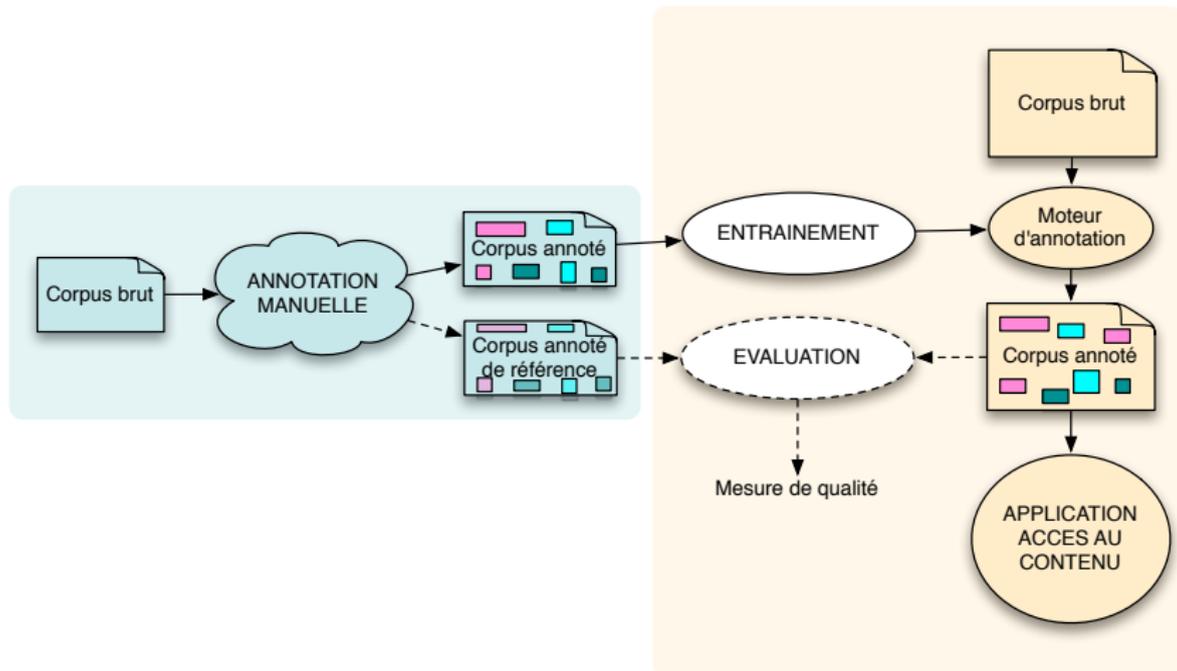
Corpus annotés et Traitement Automatique des Langues



Corpus annotés et Traitement Automatique des Langues



Corpus annotés et Traitement Automatique des Langues



Produire des corpus annotés manuellement : l'enjeu du coût

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

ESTER

- ▶ 100 h de parole transcrite (campagne d'évaluation ESTER, systèmes de transcription, 2008)
- ▶ 1 h de parole = entre 20 et 60 h de travail de transcription

Produire des corpus annotés manuellement : l'enjeu du coût

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

ESTER

- ▶ 100 h de parole transcrite (campagne d'évaluation ESTER, systèmes de transcription, 2008)
- ▶ 1 h de parole = entre 20 et 60 h de travail de transcription

→ Nécessité de trouver des solutions pour réduire ce coût

La myriadisation (*crowdsourcing*)

[Sagot et al., 2011] ou *externalisation ouverte*, ou *production participative*

***Crowdsourcing** is "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."*

[Howe, 2006]

- ▶ **pas d'identification** ou de sélection *a priori* des participants ("open call")

La myriadisation (*crowdsourcing*)

[Sagot et al., 2011] ou *externalisation ouverte*, ou *production participative*

Crowdsourcing is "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."

[Howe, 2006]

- ▶ **pas d'identification** ou de sélection *a priori* des participants ("open call")
- ▶ **massive** (dans la production et la participation)

La myriadisation (*crowdsourcing*)

[Sagot et al., 2011] ou *externalisation ouverte*, ou *production participative*

Crowdsourcing is "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."

[Howe, 2006]

- ▶ **pas d'identification** ou de sélection *a priori* des participants ("open call")
- ▶ **massive** (dans la production et la participation)
- ▶ (relativement) **peu coûteux** (répond à la préoccupation du coût)

Une taxinomie (simplifiée) de la myriadisation

parmi d'autres, voir [Geiger et al., 2011]

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

Une taxinomie (simplifiée) de la myriadisation

parmi d'autres, voir [Geiger et al., 2011]

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

Traitement Automatique des Langues (TAL) : l'ogre des données

Amazon Mechanical Turk : la solution ou le problème ?

Pour la petite histoire

Page d'accueil

AMT : la construction du mythe

TATA (« There Are Thousands of Alternatives »)

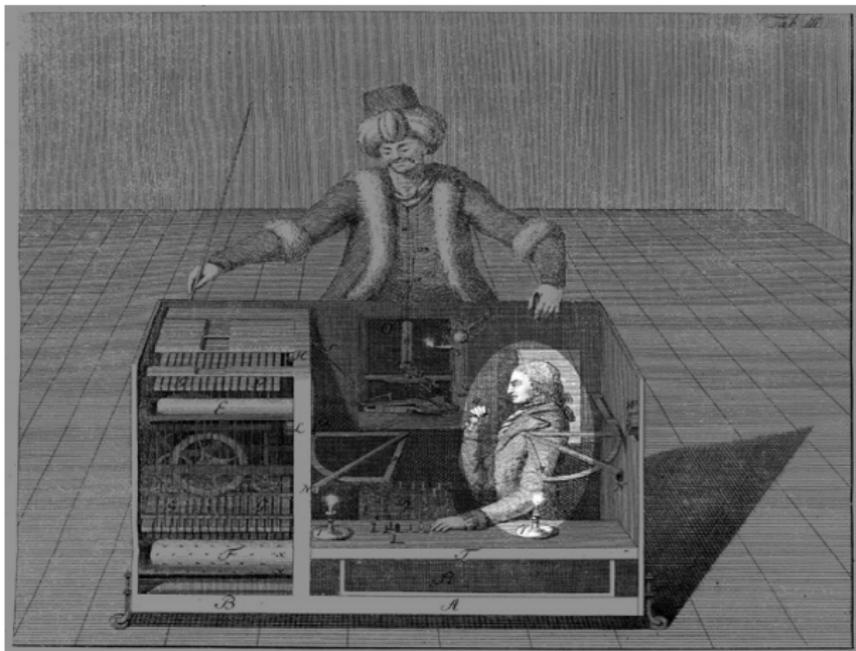
Le «Turc mécanique» de von Kempelen

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



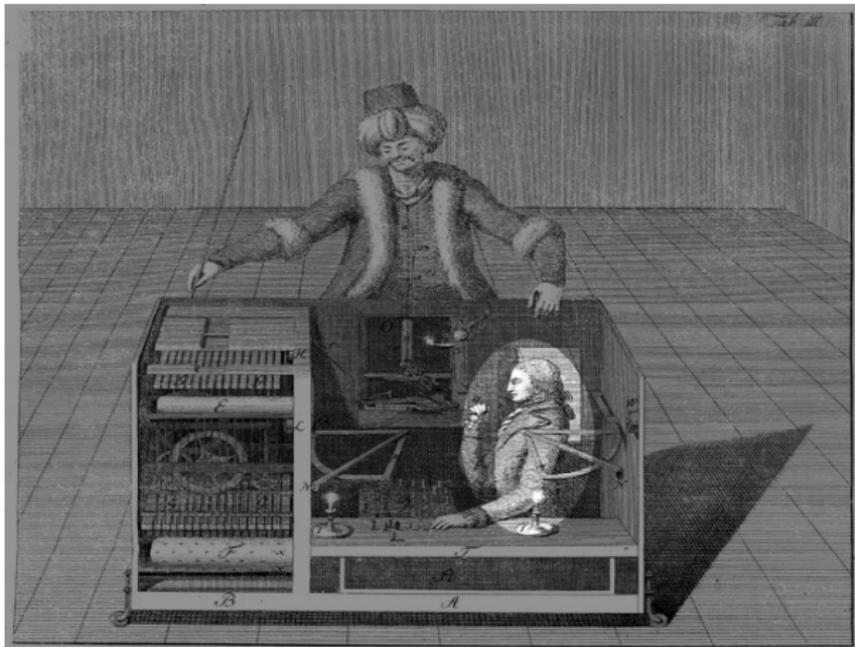
Le «Turc mécanique» de von Kempelen

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



Le «Turc mécanique» de von Kempelen

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

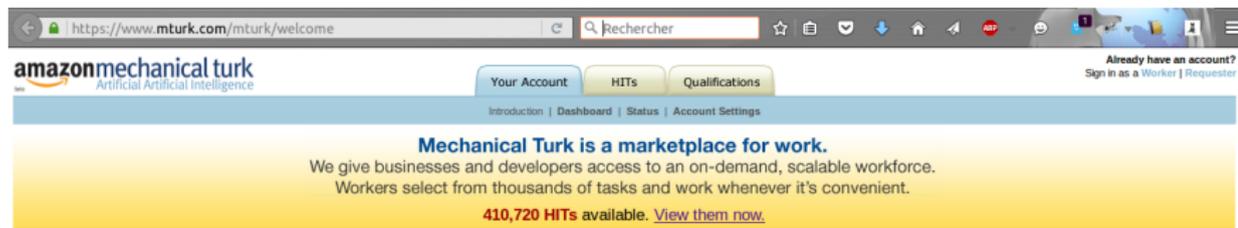


Et Amazon créa AMT

Amazon crée pour ses propres besoins de catégorisation une
plate-forme de travail parcellisé
et en ouvre l'accès en 2005
(moyennant, à l'origine, 10 % des transactions)

Amazon Mechanical Turk (AMT)

AMT



https://www.mturk.com/mturk/welcome

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.
410,720 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](#)

or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

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

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 (AMT)

AMT est une plate-forme de **myriadisation** : le travail est **externalisé** via le Web et réalisé par des internautes, les **Turkers**

The screenshot shows the Amazon Mechanical Turk homepage. At the top, there is a search bar with the text "Rechercher" and a navigation menu with "Your Account", "HITS", and "Qualifications". Below the navigation, there is a banner with the text "Mechanical Turk is a marketplace for work." and "We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient." Below the banner, it says "410,720 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



or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

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

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



Amazon Mechanical Turk (AMT)

AMT 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**

The screenshot shows the Amazon Mechanical Turk website interface. At the top, there is a navigation bar with the Amazon Mechanical Turk logo and the text 'Artificial Intelligence'. Below the logo, there are buttons for 'Your Account', 'HITS', and 'Qualifications'. The main content area features a yellow banner with the text: '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. 410,720 HITs available. View them now.' Below this banner, there are two main sections: 'Make Money by working on HITs' and 'Get Results from Mechanical Turk Workers'. The 'Make Money' section includes a green-bordered box with the text: 'HITs - Human Intelligence Tasks - are individual tasks that you work on. Find HITs now.' Below this, it lists benefits for workers: 'Can work from home', 'Choose your own work hours', and 'Get paid for doing good work'. A diagram shows a flow from 'Find an interesting task' to 'Work' to 'Earn money'. The 'Get Results' section includes a list of benefits for requesters: 'Have access to a global, on-demand, 24 x 7 workforce', 'Get thousands of HITs completed in minutes', and 'Pay only when you're satisfied with the results'. A diagram shows a flow from 'Fund your account' to 'Load your tasks' to 'Get results'.

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](#)

or [learn more about being a Worker](#)

Get Results
from Mechanical Turk Workers

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

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 (AMT)

AMT 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's a search bar with the text "Rechercher". Below it, the Amazon Mechanical Turk logo is displayed. Navigation buttons for "Your Account", "HITS", and "Qualifications" are visible. A banner below the navigation bar states: "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. 410,720 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



or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

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

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



Amazon Mechanical Turk (AMT)

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

The screenshot shows the Amazon Mechanical Turk homepage. At the top, there is a search bar with the text "Rechercher" and a navigation menu with "Your Account", "HITS", and "Qualifications". Below the navigation, a banner reads: "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. 410,720 HITS available. View them now." The page is designed with a clean, professional layout using blue and yellow accents.

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



or [learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

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

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**



Caractéristiques d'AMT

Rémunération :

- ▶ à la (micro) tâche (*illégal* en France sauf (rares) exceptions) : moins de 2 \$/h (70 % des HIT sont payées moins de 0,05 \$ US [Ipeirotis, 2010b])
- ▶ pas de relation explicite entre les *Turkers* et les *Requesters*

Tâches :

- ▶ nouveaux usages : par exemple, des créations artistiques, comme <http://www.thesheepmarket.com/>
- ▶ des tâches traditionnellement réalisées par des employés salariés : transcription, traduction (agences LDC, ELDA), etc

Types de tâches les plus courants sur AMT

<http://www.mturk-tracker.com>, février 2016



General

Top-1000 Requesters, report for February 3, 2016 to March 4, 2016

Arrivals

Completions

Top requesters

Active requesters

Demographics

Search

API

Old version

Requester name	hits	reward
Jon Breilig	84646	\$12,939.88
SET Master Account	75842	\$816.35
VDX	68792	\$1,634.16
p9r	58274	\$2,199.26
Amazon Requester Inc - browse classification	40297	\$2,429.52
AI Indoors Project	26934	\$9,669.04
OCMP5	26579	\$1,232.27
NoblisCV	16337	\$7,122.20
PDS_PN1	14808	\$406.49
HITmeUp	14374	\$2,872.14

< Previous

10 25 50 100 Page 1 of 100

Next >

- ▶ identifications diverses sur des tickets de caisse
- ▶ classification binaire d'images (est/n'est pas sur l'image)
- ▶ ... et transcription (*CastingWords*), dialogue (*Dialogue Systems*), etc.

AMT : le rêve devenu réalité ?

pour le Traitement Automatique des Langues

Cheap and Fast — But is it Good? **Evaluating Non-Expert Annotations for Natural Language Tasks**

Rion Snow[†] Brendan O'Connor[‡] Daniel Jurafsky[§] Andrew Y. Ng[†]

[†]Computer Science Dept.
Stanford University
Stanford, CA 94305
{rion,ang}@cs.stanford.edu

[‡]Dolores Labs, Inc.
832 Capp St.
San Francisco, CA 94110
brendano@doloreslabs.com

[§]Linguistics Dept.
Stanford University
Stanford, CA 94305
jurafsky@stanford.edu

[Snow et al., 2008]

AMT : le rêve devenu réalité ?

pour le Traitement Automatique des Langues

Cheap and Fast — But is it Good? **Evaluating Non-Expert Annotations for Natural Language Tasks**

Rion Snow[†] Brendan O'Connor[‡] Daniel Jurafsky[§] Andrew Y. Ng[†]

[†]Computer Science Dept.
Stanford University
Stanford, CA 94305
{rion,ang}@cs.stanford.edu

[‡]Dolores Labs, Inc.
832 Capp St.
San Francisco, CA 94110
brendano@doloreslabs.com

[§]Linguistics Dept.
Stanford University
Stanford, CA 94305
jurafsky@stanford.edu

[Snow et al., 2008]

C'est très peu cher, rapide, de bonne qualité
et c'est un hobby pour les *Turkers* !

AMT permet de réduire les coûts

Très basse rémunération \Rightarrow coûts faibles ? Oui, mais. . .

- ▶ coût de mise au point de l'**interface**
- ▶ coût de création de protections contre les **spammers**
- ▶ coût de **validation** et de **post-traitement**

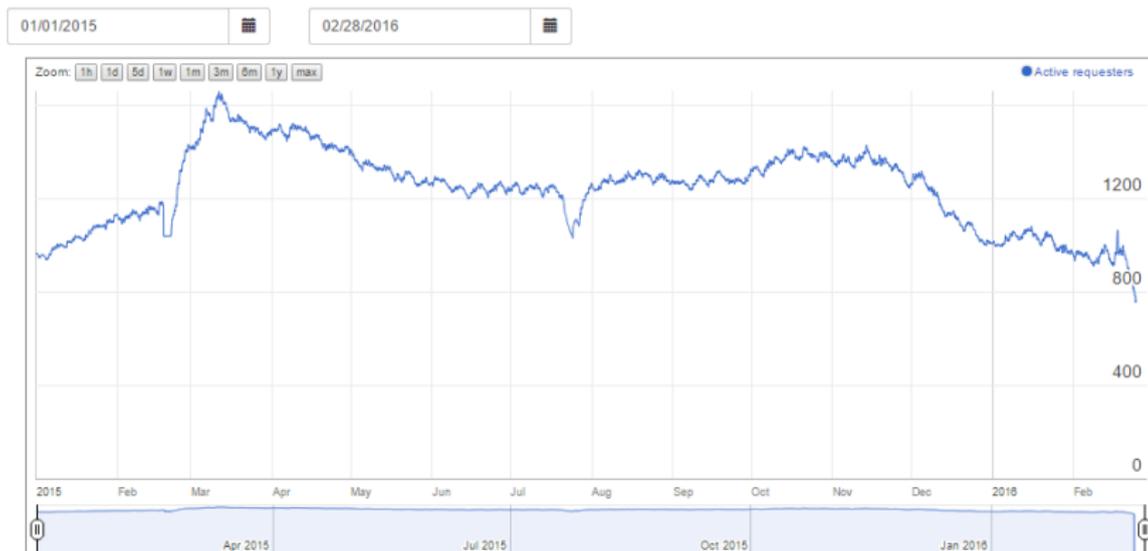
certaines tâches (par exemple, la traduction du coréen vers l'anglais) génèrent des coûts similaires aux coûts habituels dans le domaine, du fait du **manque de *Turkers* qualifiés** [Novotney and Callison-Burch, 2010].

Quand Amazon se sert...

Amazon is doubling the fee it collects from "requesters," those seeking laborers to perform online tasks, to 20% beginning July 21. And for tasks requiring at least 10 people, Amazon will charge an additional 20%, a new fee.

[blog du Wall Street Journal, 23 juin 2015]

Conséquences (?)



AMT permet de produire des ressources de qualité ?

- ▶ permet de produire des ressources de qualité dans certains cas précis (par exemple, la transcription simple)
- ▶ mais :
 - ▶ la qualité est insuffisante lorsque la tâche est **complexe** (par exemple, le résumé [Gillick and Liu, 2010])
 - ▶ l'**interface** d'AMT pose parfois problème [Tratz and Hovy, 2010]
 - ▶ les *Turkers* posent parfois problème (tricheurs, **spammers**)
 - ▶ le modèle de rémunération **à la tâche** pose problème [Kochhar et al., 2010]
- ▶ pour certaines tâches simples les outils de TAL produisent de **meilleurs résultats** qu'AMT [Wais et al., 2010].

AMT : un passe-temps pour les *Turkers* ?

[Ross et al., 2010, Ipeirotis, 2010a] montrent que :

- ▶ les *Turkers* sont avant tout motivés par l'**argent** (91 %) :
 - ▶ 20 % considèrent AMT 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 AMT, et contribuent à 80 % des tâches
- ▶ le salaire horaire moyen observé est **inférieur à 2 \$**

[Gupta et al., 2014] : en l'absence de possibilité de formation, un important **travail caché** est réalisé par les *Turkers*

500 000 *Turkers* ?

- ▶ 500 000 *Turkers* selon AMT
- ▶ en réalité, entre 15 059 et 42 912 [Fort et al., 2011]

Est-ce qu'AMT est éthique et/ou légal ?

Éthique :

- ▶ pas d'**identification** : pas de lien officiel entre *Requesters* et *Turkers* et entre *Turkers*
- ▶ pas de possibilité de **se syndiquer**, pour protester contre des manquements des *Requesters* ou ester en justice
- ▶ pas de **salaire minimum** (< 2 \$/h en moyenne)
- ▶ possibilité de **refuser de payer** les *Turkers*

Est-ce qu'AMT est éthique et/ou légal ?

The screenshot shows the Amazon Mechanical Turk website. At the top, there is a navigation bar with the Amazon Mechanical Turk logo and the text "Artificial Intelligence". To the right, there are buttons for "Your Account", "HITS", and "Qualifications". Below this, there is a navigation menu with "Introduction", "Dashboard", "Status", and "Account Settings". The main content area features a yellow banner with the text: "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. 410,720 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



[or learn more about being a Worker](#)

Get Results from Mechanical Turk Workers

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

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**



Est-ce qu'AMT est éthique et/ou légal ?

Légalité :

- ▶ accord de licence d'Amazon : les *Turkers* sont considérés comme des travailleurs indépendants ⇒ ils sont supposés se déclarer comme tels et payer les cotisations afférentes
- ▶ illusoire, vu le niveau de rémunération
- ⇒ les États **perdent** une source de revenus légitime

Ce qu'AMT fait au TAL : TINA (?)

(There Is No Alternative)

- ▶ *crowdsourcing* est devenu synonyme d'AMT
- ▶ utilisation qui se généralise en TAL anglophone (et pas seulement) :
 - ▶ 3 articles à ACL 2010, 16 en 2015
 - ▶ 5 articles à COLING 2010, 10 en 2014
- ▶ **simplification** des tâches
 - ▶ une « vraie » tâche d'annotation en inférences textuelles (inférence, neutre, contradiction) est réduite à 2 phrases et une question :
« Would most people say that if the first sentence is true, then the second sentence must be true ? » [Bowman et al., 2015]
- ▶ difficile d'argumenter pour de « vrais » **financements**

Ce qu'AMT fait au TAL : prise de conscience

- ▶ déconstruire les **mythes** [Fort et al., 2011, Fort et al., 2014a]
- ▶ formaliser des **bonnes pratiques** [Adda et al., 2013]
- ▶ créer des outils pour aider les *Turkers* à mieux gagner leur vie [Callison-Burch, 2014]
*[mais nous avons mené une expérience qui montre que les *Turkers*, quand ils sont mieux payés, mettent plus de temps à réaliser les tâches (!)]*
- ▶ mieux comprendre la myriadisation
- ▶ poser les questions d'**éthique** dans le TAL

Traitement Automatique des Langues (TAL) : l'ogre des données

Amazon Mechanical Turk : la solution ou le problème ?

TATA (« There Are Thousands of Alternatives »)

Autres formes de myriadisation

Éthique et TAL

Autres formes de myriadisation

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

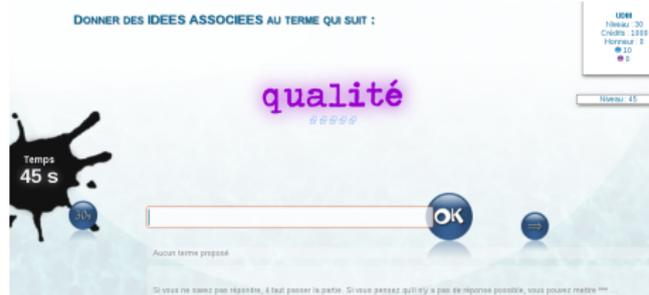
JeuxDeMots : jouer à associer des idées...

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

Plus de **10 millions de relations** créées et **constamment** mises à jour

Mécanisme très élaboré :

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



→ Profiter des **capacités** de la foule

ZombiLingo : jouer à manger des têtes ...

... pour annoter des corpus en syntaxe de dépendances [Fort et al., 2014b]

Jouer Boutique Forum Admin Joueurs

Trouve la tête de ce qui est introduit par la préposition indiquée

0%

Prochain niveau
477882/620000
304

Besoin d'aide ?

Pour les SCA, la durée **DE** la perfusion dépend de la manière dont le SCA doit être traité : elle peut durer jusqu' à 72 heures au maximum chez les patients devant recevoir des médicaments.

58

49 150 Acheter
11 15 Acheter
2 300 Acheter
2 15 Acheter

- ▶ décomposition de la tâche par phénomènes (vs simplification)
- ▶ formation des joueurs (au début, puis vérification régulière)

→ Profiter des **capacités d'apprentissage** de la foule

Retour d'expérience : ZombiLingo

début mars 2016

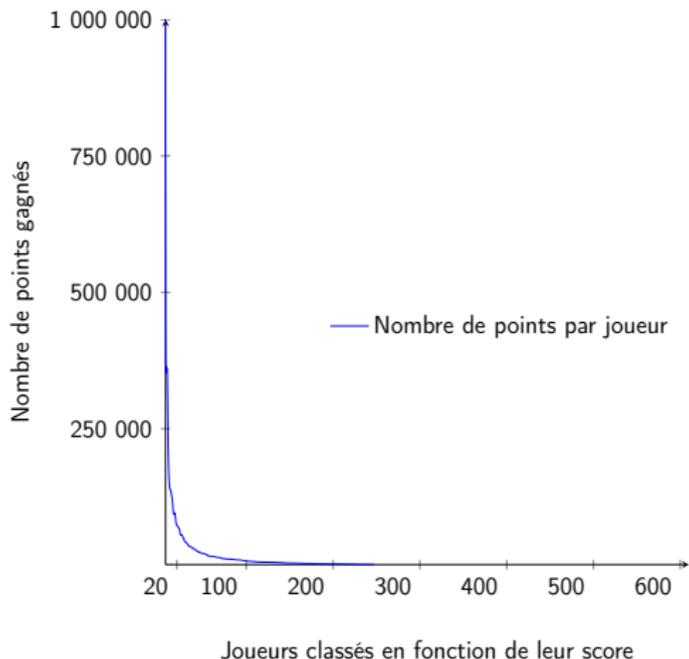
Quantitativement :

- ▶ 608 inscrits
- ▶ 10 joueurs ont réalisé plus de 2 500 annotations
- ▶ semaines à plus de 10 000 annotations
- ▶ plus de 97 000 annotations réalisées en tout

Qualitativement :

- ▶ 86 % d'exactitude sur la référence pour les 10 plus gros joueurs
- ▶ près de 90 % si on inclut tous les joueurs
- ▶ mais surévaluée (génération de bruit sur les phrases pas de référence) ⇒ à améliorer

ZombiLingo : nb joueurs en fonction de leur classement



Redéfinir la myriadisation

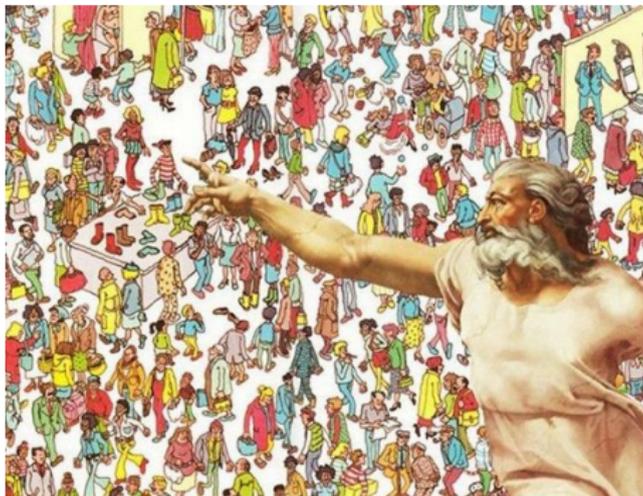
Production par des « non-experts » ?



Redéfinir la myriadisation

~~Production par des « non-experts » ?~~

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



Myriadisation éthique : perspectives ?

Une plate-forme nationale de **sciences participatives** visant à mutualiser les efforts de :

- ▶ communication
- ▶ gestion de la communauté
- ▶ développement
- ▶ hébergement (serveurs)

tout en proposant une **réflexion éthique**

Éthique et TAL : se réunir, réfléchir, agir

- ▶ Charte Éthique et Big Data :
<http://wiki.ethique-big-data.org/>
 - ▶ autodéclarée
 - ▶ 3 volets : traçabilité, propriété intellectuelle, questions juridiques
 - ▶ identifie tous les acteurs et leur mode de rémunération
- ▶ Journées d'études :
 - ▶ Journée d'études ATALA Éthique et TAL, nov. 2014
 - ▶ Atelier TALN ETeRNAL, juin 2015
 - ▶ Journée Éthique et données, jan. 2016
- ▶ Numéro spécial de la revue TAL
- ▶ Blog éthique et TAL <http://www.ethique-et-tal.org/>

Merci

Merci de jouer !



<http://zombilingo.org>

Annexes

Compléments sur AMT

Foule Factory

Plus gros Requesters (en \$ dépensés)

<http://www.mturk-tracker.com>, février 2016



[General](#)

[Arrivals](#)

[Completions](#)

[Top requesters](#)

[Active requesters](#)

[Demographics](#)

[Search](#)

[API](#)

[Old version](#)

Top-1000 Requesters, report for February 3, 2016 to March 4, 2016

Requester name	hits	reward
Colleen Harmeling	1451	\$50,725.82
Jon Breilig	84646	\$12,939.88
CastingWords	12029	\$10,499.14
AI Indoors Project	26934	\$9,669.04
NobilisCV	16337	\$7,122.20
Groupsolver	3359	\$6,563.45
HITmeUp	14374	\$2,872.14
Amazon Requester Inc - browse classification	40297	\$2,429.52
p9r	58274	\$2,199.26
Dialogue Systems	6899	\$1,935.55

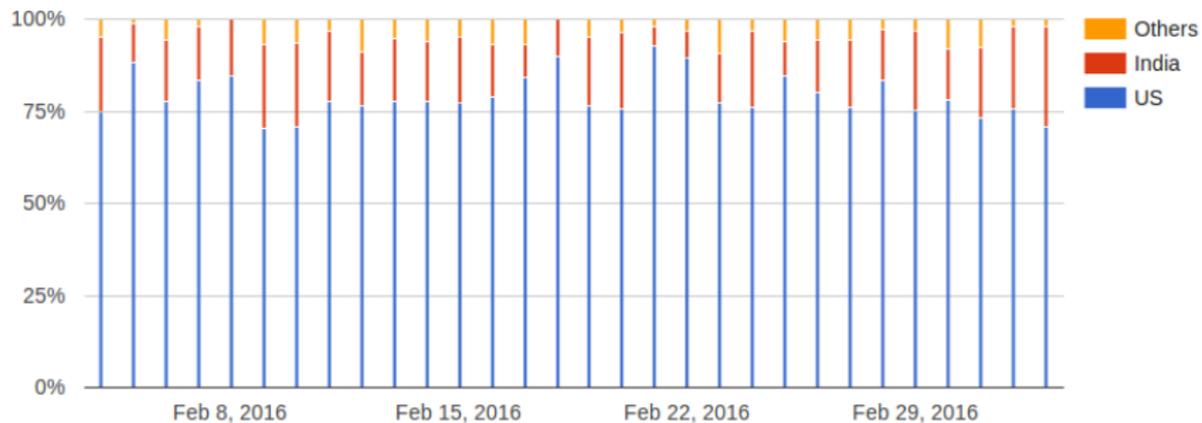
[« Previous](#)

[10](#) [25](#) [50](#) [100](#) Page 1 of 100

[Next »](#)

Provenance des *Turkers*

<http://www.mturk-tracker.com>, février 2016



Foule Factory : version plus éthique d'AMT, à la française ?

«[...] nous avons (contrairement à Amazon) mis en place plusieurs leviers pour s'assurer d'une certaine éthique dont l'auteure pense qu'elle nous fait défaut. a/ Foulefactory n'est ouvert qu'aux contributeurs français. b/ Le plafond annuel de rémunération est de 3 000 euros - nous ne souhaitons pas que foulefactory constitue une source primaire de revenus pour nos contributeurs. c/ Le taux horaire moyen est compris entre 10 et 15 euros, c'est le prix minimum que nous demandons à nos clients de respecter. d/ Si un projet rémunère moins, nos fouleurs (contributeurs de foulefactory) nous remontent l'information et nous adressons un avertissement au client. e/ Si une tâche réalisée par un fouleur est refusée (et donc non-rémunérée), ce dernier peut faire une « réclamation » en 1 clic. f/ Si plusieurs fouleurs adressent une réclamation à un client, nous adressons un avertissement à ce dernier. g/ Après 3 avertissements, un client est banni de la plateforme foulefactory.[...] »

[extrait du commentaire laissé par D. Benoïlid co-fondateur de FF, sur mon [article](#) Journal du CNRS / Libération sur la myriadisation]

« c/ Le taux horaire moyen est compris entre 10 et 15 euros, c'est le prix minimum que nous demandons à nos clients de respecter. »



Mel Anie 10/15 euros de l'heure... n'importe quoi... très très rares sont les projets qui atteignent ce montant....

J'aime · Répondre · 29 juillet 2015, 14:02



Dominique Bonhomme oui mais tu regardes l'info, ils sont dans la légalité, dans la mesure où il mesure ta tâche 10 s

J'aime · Répondre · 29 juillet 2015, 16:13



Mel Anie la plupart du temps, les projets sont plutôt autour de 7-9 euros de l'heure

J'aime · Répondre · 29 juillet 2015, 16:15 · Modifié



foulefactory

5 0 2 days ago

Identifiant: #5326

Titre: Normalization 190220161709

Catégorie: Rédaction

Temps par tâche: 00:00:05

Défraiment: 0.01 €

Taux horaire: 7.2 €

Nombre de Fouleur: 1

Date de creation: 2/19/2016 5:12:48 PM

Indices extraits du forum <http://www.fouleurs.com/>

« f/ Si plusieurs fouleurs adressent une réclamation à un client, nous adressons un avertissement à ce dernier. »

lipqil 9 0 on 9 février 2016

Pour moi aussi, problème sur le 5294. Dans « Mes réalisations » j'ai 326 tâches validées (pour 2.46€), alors que dans la détail du projet il a de mentionné: **Validées: 41, Refusées: 0, En attente: 0**. Alors ont sont passées les 285 tâches restantes. C'EST DE LA PURE ARNAQUE!!

Pour le projet 5234 les tâches sont en attente, la date de fin est 30/11/2017, On peut toujours espérer. Il n'y aura pas de paiement.

Et comme toujours, aucune réponse de FF sur les multiples explications que je leur ai envoyées sur de nombreux problèmes.

 **Samuel** 0 on 9 février 2016

J'ai contacté FF j'attends qu'il me réponde.

lipqil 0 on 9 février 2016

C'est bien, mais il ne faut pas espérer avoir une réponse. Qui a eu des réponses de leur part ces dernières semaines?

Moi, aucune depuis des mois.

 **Samuel** 0 on 9 février 2016

je vous redit

 **Samuel** 0 on 9 février 2016

quand j'aurais une réponse

lipqil 0 on 9 février 2016

Oui, oui. J'entends bien. Comme on dit, l'espoir fait vivre

Foule Factory : dépassé par son succès ?

Certains des problèmes sont sans doute dus à l'afflux soudain et massif de travailleurs suite à une émission de TV.

Cependant :

- ▶ Déficit de tâches à réaliser : du fait de la loi de l'offre et de la demande, il paraît peu probable que les déposants soient inquiétés...
- ▶ Taux horaire défini comment ?

-  Adda, G., Mariani, J.-J., Besacier, L., and Gelas, H. (2013). Economic and ethical background of crowdsourcing for speech. In Crowdsourcing for Speech Processing : Applications to Data Collection, Transcription and Assessment, pages 303–334. Wiley.
-  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.
-  Bowman, S. R., Angeli, G., Potts, C., and Manning, C. D. (2015). A large annotated corpus for learning natural language inference. arXiv preprint arXiv :1508.05326.
-  Callison-Burch, C. (2014).

Crowd-workers : Aggregating information across turkers to help them find higher paying work.

In The Second AAAI Conference on Human Computation and Crowdsourcing (HCOMP-2014).



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

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



Fort, K., Adda, G., Sagot, B., Mariani, J., and Couillault, A. (2014a).

Crowdsourcing for Language Resource Development :
Criticisms About Amazon Mechanical Turk Overpowering Use.

In Vetulani, Z. and Mariani, J., editors,
Human Language Technology Challenges for Computer Science and L
pages 303–314. Springer International Publishing.



Fort, K., Guillaume, B., and Stern, V. (2014b).

Zombilingo : eating heads to perform dependency syntax annotation (zombilingo : manger des têtes pour annoter en syntaxe de dépendances) [in french].

In Proceedings of TALN 2014 (Volume 3 : System Demonstrations), pages 15–16, Marseille, France. Association pour le Traitement Automatique des Langues.



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.



Gillick, D. and Liu, Y. (2010).

Non-expert evaluation of summarization systems is risky.

In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, CSLDAMT '10, pages 148–151, Stroudsburg, PA, USA. Association for Computational Linguistics.

 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.

 Gupta, N., Martin, D., Hanrahan, B. V., and O'Neill, J. (2014).

Turk-life in india.

In Proceedings of the 18th International Conference on Supporting Group Work, GROUP '14, pages 1–11, New York, NY, USA. ACM.

 Howe, J. (2006).

The rise of crowdsourcing.

Wired Magazine, 14(6).

 Ipeirotis, P. (2010a).

The new demographics of mechanical turk.

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



Ipeirotis, P. G. (2010b).

Analyzing the amazon mechanical turk marketplace.

XRDS, 17(2) :16–21.



Kochhar, S., Mazzocchi, S., and Paritosh, P. (2010).

The anatomy of a large-scale human computation engine.

In Proceedings of Human Computation Workshop at the 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, KDD 2010, Washington D.C.



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.



Latour, B. (1987).

Science in Action : How to Follow Scientists and Engineers through Society.

Harvard University Press, Cambridge, MA.



Novotney, S. and Callison-Burch, C. (2010).

Cheap, fast and good enough : automatic speech recognition with non-expert transcription.

In Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), HLT'10, pages 207–215, Stroudsburg, PA, USA. Association for Computational Linguistics.



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.



Sagot, B. (2010).

The Lefff, a freely available and large-coverage morphological and syntactic lexicon for French.

In

7th international conference on Language Resources and Evaluation (Valletta, Malta.



Sagot, B., Fort, K., Adda, G., Mariani, J., and Lang, B. (2011).

Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé.

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

12 pages.



Snow, R., O'Connor, B., Jurafsky, D., and Ng., A. Y. (2008).

Cheap and fast - but is it good? evaluating non-expert annotations for natural language tasks.

In Proceedings of EMNLP 2008, pages 254–263.



Tratz, S. and Hovy, E. (2010).

A taxonomy, dataset, and classifier for automatic noun compound interpretation.

In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 678–687, Uppsala, Suède. Association for Computational Linguistics.



Wais, P., Lingamneni, S., Cook, D., Fennell, J., Goldenberg, B., Lubarov, D., Marin, D., and Simons, H. (2010).

Towards building a high-quality workforce with mechanical turk. In Proceedings of Computational Social Science and the Wisdom of Crowds (NIPS).