



**HAL**  
open science

## La production participative (crowdsourcing) : miroir grossissant sur l'annotation manuelle

Karën Fort

► **To cite this version:**

Karën Fort. La production participative (crowdsourcing) : miroir grossissant sur l'annotation manuelle. Séminaire de l'équipe Inria ALMANACH, Oct 2019, Paris, France. hal-02418852

**HAL Id: hal-02418852**

**<https://inria.hal.science/hal-02418852>**

Submitted on 19 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# La production participative (*crowdsourcing*) : miroir grossissant sur l'annotation manuelle

Karën Fort

karen.fort@sorbonne-universite.fr

11 octobre 2019





## Production participative (*crowdsourcing*)

Jeux ayant un but que j'ai participé à créer :



ZOMBLINGO

RIGORMORTIS

BISAME

KRIK!

AYO!



Portail de jeux pour les langues et atelier récurrent :

L I N G O B O I N G O

Games4NLP

L'annotation manuelle de corpus

La production participative

ZombiLingo

Conclusion

## L'annotation manuelle de corpus

Annoter c'est interpréter

Une activité (insuffisamment) outillée

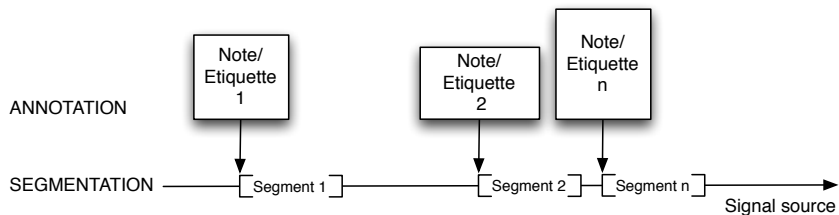
Annotateurs et biais

La production participative

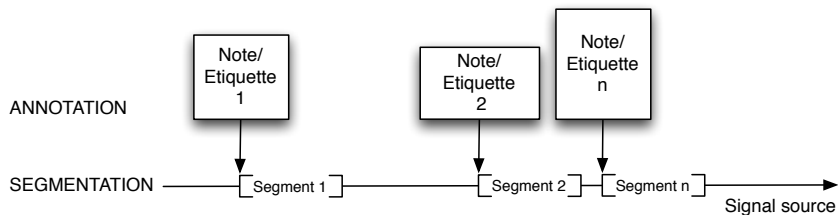
ZombiLingo

Conclusion

# Définition de l'annotation



# Définition de l'annotation



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



# La science de la catégorisation

## Mesurer vs quantifier

Certaines réalités sont immédiatement mesurables :

- ▶ la hauteur du Mont Everest (8 848 m)

d'autres non :

- ▶ le nombre de chômeurs → qu'est-ce qu'un chômeur ?

"Mais précisément la définition et la mesure de la population active et du chômage relèvent d'une autre épistémologie que celle de l'étoile polaire. Elles impliquent des conventions (analogues aux principes généraux des lois et des codes votés par les Parlements) et des décisions (analogues à celles d'un juge) d'affecter tel cas à telle classe." [Desrosières, 2001]

# Coder

” Un codage est une décision conventionnelle de construire une classe d'équivalence entre divers objets, la 'classe' étant jugée plus 'générale' que tout objet singulier. La 1ère condition pour cela est de supposer que tous ces objets peuvent être comparés, ce qui ne va pas de soi”

[Desrosières, 1989]

→ Quantifier, c'est [se mettre d'accord](#), puis mesurer

[Desrosières, 2008]

# Comment se mettre d'accord ?

guide d'annotation

et comment vérifier que le consensus est compris et appliqué ?

accord inter-annotateur

## L'annotation manuelle de corpus

Annoter c'est interpréter

Une activité (insuffisamment) outillée

Annotateurs et biais

La production participative

ZombiLingo

Conclusion

# Des outils d'aide à l'annotation...

WebAnno : <https://webanno.github.io>

The screenshot displays the WebAnno web application interface. At the top, a red header bar contains the text "Annotation" and "Home". To the right of the header, there are links for "Help", "admin", and "Log out (automatically in 29 min)".

Below the header is a toolbar with five main sections: "Document", "Page", "Script", "Help", and "Workflow".

- Document:** Includes icons for Open, Prev, Next, Export, and Settings.
- Page:** Includes navigation icons for First, Prev, Go to (with "96" displayed), Next, and Last.
- Script:** Includes an icon for LTR/RTL.
- Help:** Includes an icon for Guidelines.
- Workflow:** Includes icons for Reset and Finish.

Below the toolbar, the document content is displayed. The URL "dataset/TSTRR16037920105200001.conll" is visible on the left. On the right, it says "Showing 96-100 of 557 sentences [document 67 of 70]".

The main text area shows a list of sentences with annotations:

- 96 O mesmo pode ser dito quanto às gratificações natalinas, conforme requerido através da alínea j, da vestibular, sendo o pleito, por conseguinte, procedente.
- 97 2.8 - DO SEGURO DESEMPREGO No que diz respeito ao seguro-desemprego, não tendo a entidade patronal cumprido com as obrigações trabalhistas que lhe competiam, fica obrigada, de forma substitutiva, ao pagamento de indenização equivalente relativamente ao seguro-desemprego, nos termos dos artigos 186 e 927, ambos do novel Código Civil, aplicáveis ao Direito Laboral, tendo esta indenização natureza trabalhista.
- 98 Tal entendimento, inclusive, já ficou cristalizado pelo C. TST, através da Orientação Jurisprudencial nº 211 da SDI-1, convertida na súmula 389, através da Resolução nº 129/2005 do C. TST, nos seguintes termos: " I - Inscreve-se na competência material da Justiça do Trabalho a lide entre empregado e empregador tendo por objeto indenização pelo não-fornecimento das guias do seguro-desemprego.
- 99 (ex-OJ nº 210 - Inserida em 08.11.2000) II - O não-fornecimento pelo empregador da guia necessária para o recebimento do seguro-desemprego dá origem ao direito à indenização.
- 100 (ex-OJ nº 211 - Inserida em 08.11.2000) " Por todo o exposto, defiro o pleito em tela.

On the right side of the interface, there is a "Layer" dropdown menu set to "Surface form" and a checkbox for "Forward annotation". Below that is an "Annotation" section with the text "No annotation selected!".

At the bottom of the page, a footer contains the text: "Technische Universität Darmstadt – Computer Science Department – WebAnno -- 3.3.4 (2018-02-20 11:56:08, build dc8a4e574b1c649740d62a2a380c023fad956a97)" and a red "Warnings" icon.



# ... aux outils de gestion de l'annotation

WebAnno : <https://webanno.github.io>

**Projects Settings**  
WebAnno | Home

**Projects**

- ADN
- Demo\_M2\_2018
- GSD-108
- NLP\_2017
- POS-FR-P4
- UD\_2017
- ZLUD
- ZL\_TCOF\_UD
- ZombiLingo
- old\_foot
- wc2018
- wc2018\_dep

**Details** Users Documents Layers Tagsets Guidelines Export

**General**

**Name:** ADN

**Description:**

**Script direction:** ltr ▾

**Project Types**

- annotation
- automation
- correction

## L'annotation manuelle de corpus

Annoter c'est interpréter

Une activité (insuffisamment) outillée

Annotateurs et biais

La production participative

ZombiLingo

Conclusion



## Biais : ce que nous savons

Les annotateurs bien formés sont **moins sensibles** aux biais :

- ▶ dus à la pré-annotation [Fort and Sagot, 2010]
- ▶ dus à l'outil d'annotation [Dandapat et al., 2009]

et annotent moins « au hasard »

L'utilisation d'un guide d'annotation permet d'obtenir de meilleures annotations (si) [Nédellec et al., 2006]

# Qu'est-ce qu'un expert ?

Experts :

- ▶ du **domaine** : annotation en microbiologie (renommage de noms de gènes), football, etc.
- ▶ de la **tâche** : annotation en entités nommées structurées

... contradictions et insuffisances :

- pour des entités nommées en microbiologie, vaut-il mieux des microbiologistes formés à ce qu'est une EN ou des linguistes ayant une formation en microbiologie ?
- pour des entités nommées dans de la presse ancienne, vaut-il mieux des spécialistes des EN ou des historiens ?

L'annotation manuelle de corpus

**La production participative**

Des annotateurs qu'on ne choisit pas

Au-delà des mythes

Redéfinition

ZombiLingo

Conclusion

## Production participative / myriadisation (*crowdsourcing*)

**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 des participants *a priori* (appel ouvert)
- ▶ massif (en termes de production et de participation)
- ▶ (relativement) peu cher

## Des réussites remarquables

Wikipédia<sup>1</sup> (octobre 2019) :

- ▶ plus de **158 million d'articles** en **279** langues
- ▶ plus de 800 millions de pages de la Wikipédia française vues en septembre

Distributed Proofreaders (Gutenberg Project)<sup>2</sup> :

- ▶ près de **40 000** livres numérisés et corrigés

Numérisation des déclarations de conflits d'intérêts des élus<sup>3</sup> :

- ▶ 11 095 extraits de déclarations saisis **en moins d'une semaine**
- ▶ près de 8 000 participants

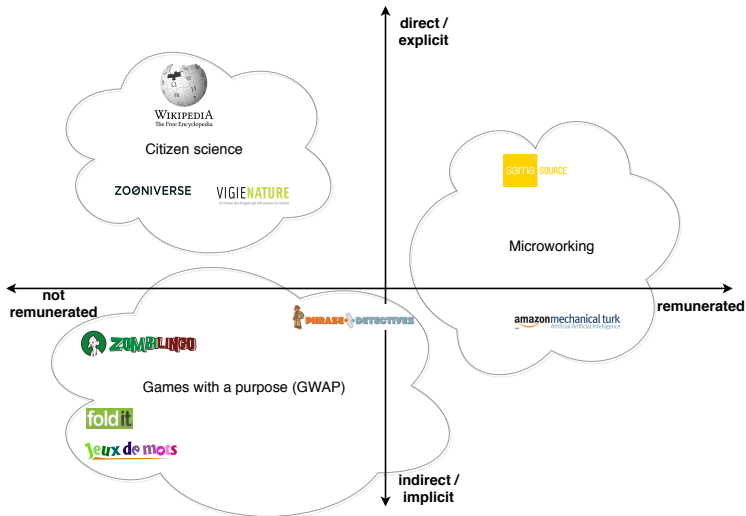
---

1. <https://stats.wikimedia.org/v2/#/all-wikipedia-projects>

2. <https://www.pgdp.net/c/>

3. <http://regardscitoyens.org/interets-des-elus/>

# Les productions participatives



L'annotation manuelle de corpus

**La production participative**

Des annotateurs qu'on ne choisit pas

**Au-delà des mythes**

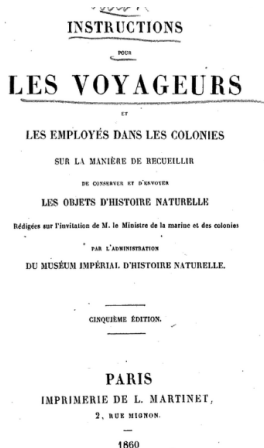
Redéfinition

ZombiLingo

Conclusion

# Mythe #1 : « C'est nouveau ! »

*Instructions pour les voyageurs et les employés des colonies*

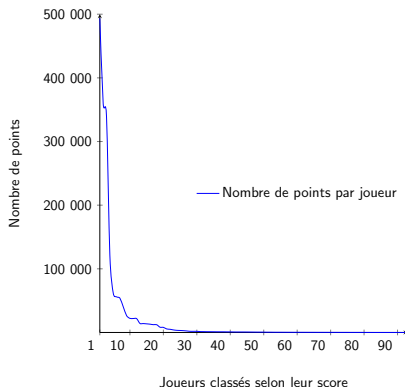


Sciences participatives :

- ▶ publiées par le Museum National d'Histoire Naturelle
- ▶ première édition en **1824**

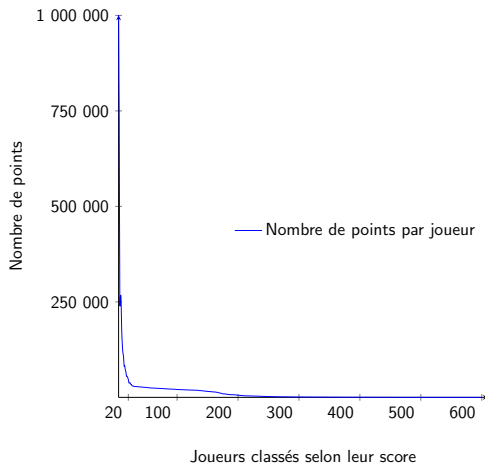


## Mythe #2 : « Ca implique beaucoup de participants »



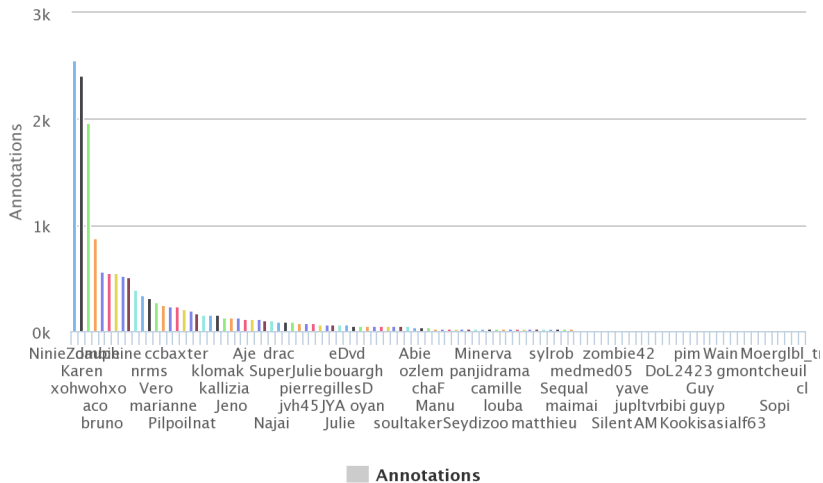
Nombre de joueurs sur *Phrase Detectives* selon le nombre de points gagnés dans le jeu (fév. 2011 - fév. 2012)

# Une foule de participants ? JeuxDeMots



Nombre de joueurs sur *JeuxDeMots* selon leur classement dans le jeu  
(source : <http://www.jeuxdemots.org/generateRanking-4.php>)

# Une foule de participants ? ZombiLingo



Highcharts.com

# Une foule de travailleurs ? [Fort et al., 2011]

Nombre de *Turkers* actifs sur Amazon Mechanical Turk :

- ▶ annoncé : plus de 500 000
- ▶ 80 % des tâches (HIT) sont réalisés par les 20 % des *Turkers* les plus actifs [Deneme, 2009]
- ⇒ travailleurs vraiment actifs (en 2011) : entre 15 059 et 42 912

# Mythe #3 : « Ca implique des non-experts »

Extraits du forum de ZombiLingo

4 zombies ont donné leur avis

1084 (Phrase de référence)



25%

Le principal critère d'évaluation de l'efficacité a été de mesurer si le taux dans le sang de phosphatases alcalines sériques ( enzyme impliquée dans la dégradation des os ) est revenu à la normale ou est **REDESCENDU** d'au moins 75 % pour se rapprocher des taux normaux .

Suivre la discussion

Discuter de la réponse

Annotation expert

Justin a écrit il y a 9 mois :

est redescendu est le passé composé du verbe redescendre, de même que "est revenu" est le passé-composé du verbe revenir. Si on les considérait comme des présents de l'indicatif au passif, il faudrait pouvoir écrire "il a été revenu" et "il a été redescendu". Ce qui n'est pas le cas...

17 zombies ont donné leur avis

22637



35%

Graphes associés à un jeu de Nim Dans la grande majorité des cas , l'ensemble des situations , même s'il est très grand , est **FINI** .

Suivre la discussion

Discuter de la réponse

Annotation expert

Zeltron a écrit il y a 9 mois :

l'ensemble est fini donc il n'est pas infini : rien à voir avec finir

L'annotation manuelle de corpus

## La production participative

Des annotateurs qu'on ne choisit pas

Au-delà des mythes

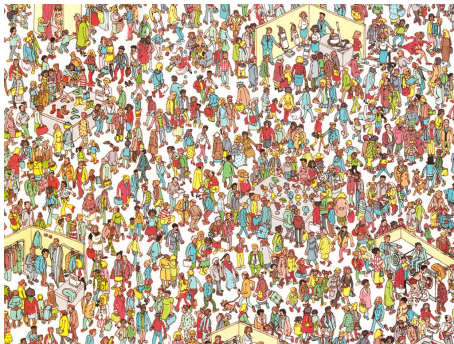
Redéfinition

ZombiLingo

Conclusion

# Production participative

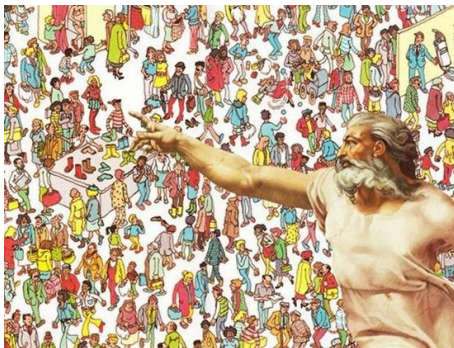
Profiter d'une foule de "non-experts" ?



# Production participative

~~Profiter d'une foule de "non-experts" ?~~

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





L'annotation manuelle de corpus

La production participative

**ZombiLingo**

La syntaxe de dépendances, en jeu

Derrière le rideau

Résultats

Conclusion

# Une tâche complexe



- ▶ guide d'annotation
  - ▶ 29 types de relation
  - ▶ approx. 50 pages
- ▶ des décisions contre-intuitives (**pas** de la grammaire d'écoliers, de la linguistique) :  $aobj = au$   
*[...] avoir recours au type de mesures [...]*  
c-à-d que la tête de la relation est ici une préposition

→ **décomposer** la complexité de la tâche [Fort et al., 2012],  
pas la simplifier !

# ZOMBI LINGO

BIENVENUE à TOI, JEUNE ZOMBIE !

LE MONDE EST  
CONDAMNÉ, TA  
TRANSFORMATION EN  
ZOMBIE A COMMENCE.

POUR SURVIVRE,  
SUIS MES RÈGLES,  
IDENTIFIE LES TÊTES  
ET MANGE-LES.

ATTENTION AUX  
PIEGES, ILS SONT  
NOMBREUX !



SIGNE

Prof. Frankenspross.

Jouer

*Pas de limite  
pour toi !  
Tu accèdes à  
toutes les  
options, bonus  
cachés !*

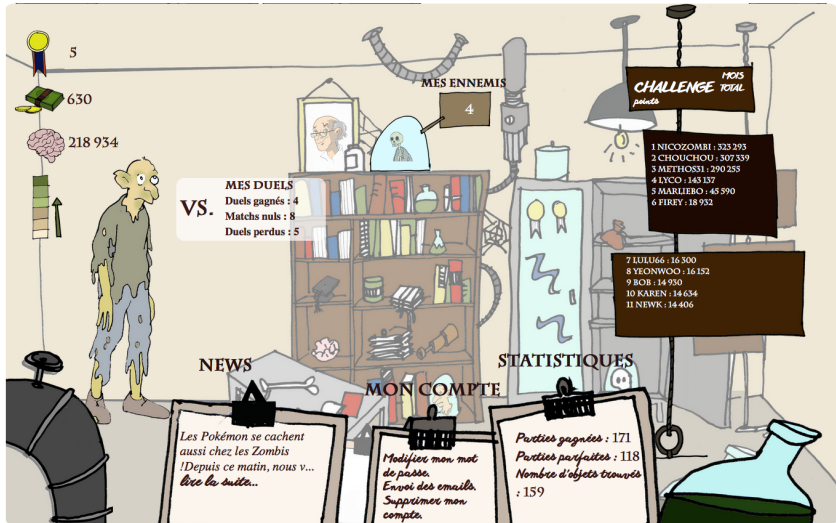
Karen

*Retrouve ici tes  
statistiques, et  
compare ton  
score avec  
celui de tes  
amis !*

SEMAINE  
MOIS  
TOTAL

1 nouveau : 38 993  
2 nicolef : 16 170

3 bruno : 11 546  
4 testKF : 7 327  
5 Karen : 7 012



AUXILIAIRE  
CAUSATIF  
Points : 290  
Annotations : 68

AUXILIAIRE DE  
TEMPS  
Points : 7 988  
Annotations : 363



CHOIX DU  
CORPUS

PERSONNE

NIVEAU 1

SUJET

JEU

FORMATION

Points : 4 980  
Annotations : 227 382

NIVEAU 1

AFFIXE

JEU

FORMATION

Points : 0  
Annotations : 0/108

NIVEAU 1

DÉTERMINANT

JEU

FORMATION

Points : 0  
Annotations : 0/148

NIVEAU 2

COMPLÈMENT EN  
« DE »

JEU

FORMATION

Points : 0  
Annotations : 0/120

NIVEAU 2

COMPLÈMENT EN  
« À »

JEU

FORMATION

Points : 0  
Annotations : 0/27

NIVEAU 2

OBJET DE LA  
PREPOSITION

JEU

FORMATION

Points : 0  
Annotations : 0/213

NIVEAU 2

AUXILIAIRE PASSIF

JEU

FORMATION

Points : 598  
Annotations : 9/56

Jouer

Boutique

Forum

Joueurs



Niveau  
maximum!

165

Trouve le complément (objet indirect introduit par "à") du verbe  
indiqué !

10%

Besoin  
d'aide?



Très jeune, il a fait preuve d'initiative et de courage pour  
**PARTICIPER** à un sauvetage lors d'inondations.



57  
150

Acheter

1  
15

Acheter

3  
300

Acheter

74

0  
15

Acheter

43

L'annotation manuelle de corpus

La production participative

## ZombiLingo

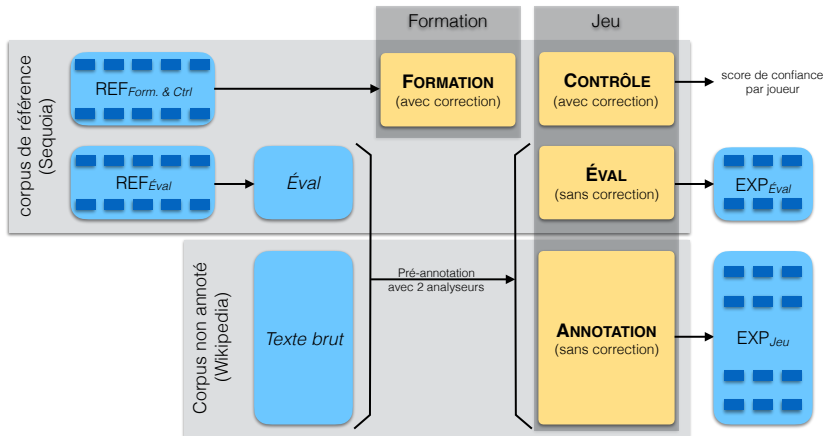
La syntaxe de dépendances, en jeu

Derrière le rideau

Résultats

Conclusion

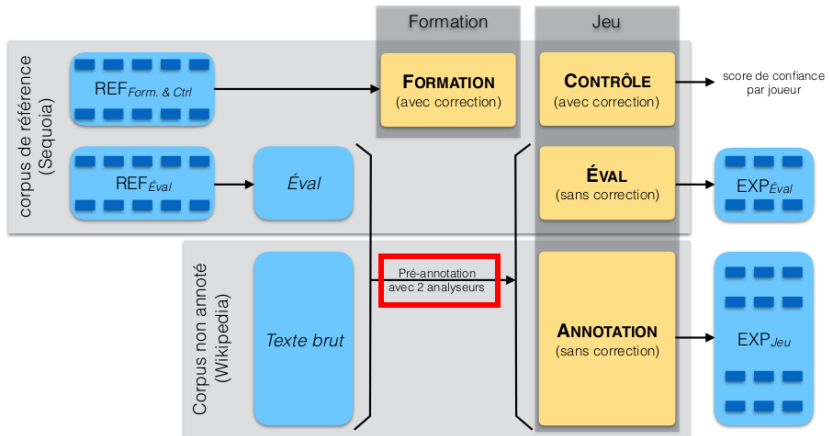
# Organiser une production de qualité





# Prétraitement des données

corpus librement disponibles et distribuables



# Prétraitement des données

corpus librement disponibles et distribuables

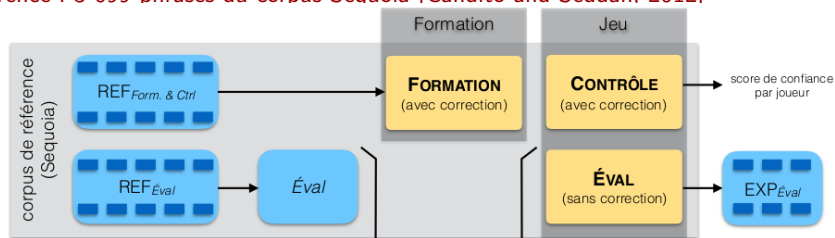
Pré-annotation avec deux parsers

1. un statistique : Talismane [Urieli, 2013]
2. un symbolique, basé sur la ré-écriture de graphes :  
FRDEP-PARSE [Guillaume and Perrier, 2015]

→ les joueurs ne jouent que les relations sur lesquelles les deux parsers ne donnent pas les mêmes résultats

# Formation, contrôle et évaluation

Référence : 3 099 phrases du corpus Sequoia [Candito and Seddah, 2012]



$REF_{Train\&Control}$	$REF_{Éval}$	non utilisé
50 %	25 %	25 %
1 549 phrases	776 phrases	774 phrases

- ▶  $REF_{Train\&Control}$  est utilisé pour former les joueurs
- ▶  $REF_{Éval}$  est utilisé comme un corpus brut, pour évaluer les annotations produites

# Formation



Obligatoire pour chaque relation

- ▶ phrases du corpus REF *Train&Control*
- ▶ retour visuel en cas d'erreur

Jouer Boutique Forum Joueurs

Besoin d'aide?

Trouve le sujet du verbe indiqué !

10%

- Septembre 2001 : le juge Eric Hanen est dessaisi de son enquête par la cour d'appel de Paris, qui ANNULE une partie de la procédure de jugement de l'arme.

Tu as répondu Paris et il fallait répondre qui

Passer à la phrase suivante

# Gestion de la fatigue cognitive et des joueurs au long court

## Mécanisme de contrôle

Des phrases de REF<sub>Train&Control</sub> sont proposées régulièrement

1. si le joueur échoue à trouver la bonne réponse, un retour visuel avec la solution lui est proposé

Ils ont été reçus à la boulangerie Leroy **POUR visiter** le fournil et **surtout** pétrir la pâte afin de confectionner de délicieux pains au chocolat qu'ils ont dégustés à l'heure du goûter avec un verre de jus de fruit.

Tu as répondu **surtout** et il fallait répondre **visiter**

Il te reste 2 essais avant de devoir refaire le tutoriel de ce phénomène

 Je ne suis pas d'accord

Passer à la phrase  
suivante



# Gestion de la fatigue cognitive et des joueurs au long court

## Mécanisme de contrôle

Des phrases de REF<sub>Train&Control</sub> sont proposées régulièrement

1. si le joueur échoue à trouver la bonne réponse, un retour visuel avec la solution lui est proposé
2. après un certain nombre d'erreurs sur une même relation, le joueur ne peut plus jouer et doit refaire la formation correspondante

- 1er FÉVRIER 1995 : Jean-Paul Schimpf, un ami intime de Didier Schuller, est arrêté sur un parking, alors que la dirigeante d'une entreprise d'assainissement disait vouloir lui remettre une somme d'argent en liquide.

Tu as répondu une et il fallait répondre arrêté

Tu as un peu oublié comment jouer ce phénomène. Pour continuer à jouer sur celui-ci, tu vas devoir refaire le tutoriel correspondant.

⚠ Je ne suis pas d'accord

Retourner au menu



# Gestion de la fatigue cognitive et des joueurs au long court

## Mécanisme de contrôle

Des phrases de REF<sub>Train&Control</sub> sont proposées régulièrement

1. si le joueur échoue à trouver la bonne réponse, un retour visuel avec la solution lui est proposé
  2. après un certain nombre d'erreurs sur une même relation, le joueur ne peut plus jouer et doit refaire la formation correspondante
- nous en déduisons un **niveau de confiance** dans le joueur, pour **cette** relation

L'annotation manuelle de corpus

La production participative

## ZombiLingo

La syntaxe de dépendances, en jeu

Derrière le rideau

Résultats

Conclusion



# Production : taille des corpus créés

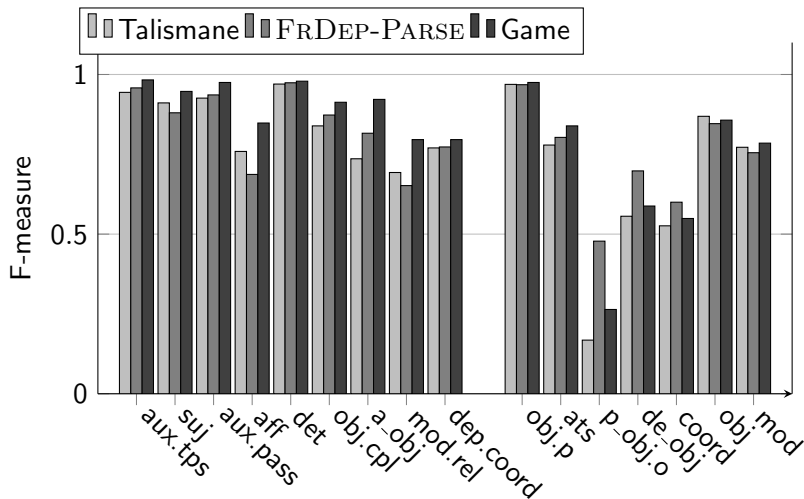
Au 10 juillet 2016

- ▶ 647 joueurs (1 460 au 9 octobre 2019)
- ▶ ont produit 107 719 annotations (496 462 au 9 octobre 2019)

→ ressource qui évolue **constamment** !

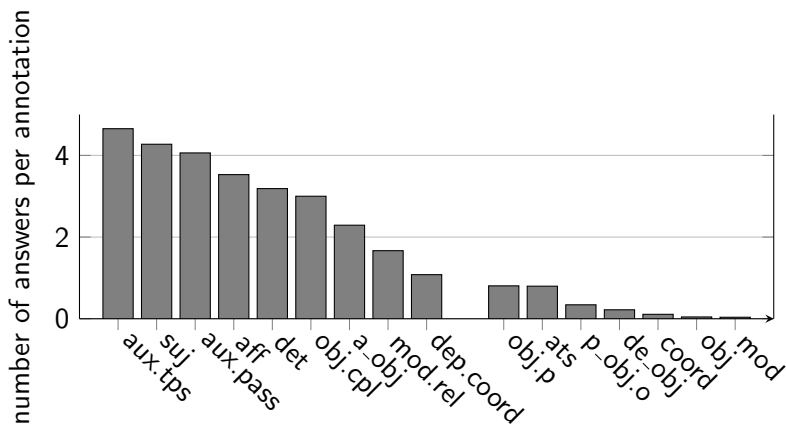
# Évaluation de la qualité

sur le corpus REF<sub>Eval</sub>



# Densité des annotations

sur le corpus REF<sub>Eval</sub>



→ besoin de **davantage** d'annotations sur certaines relations

L'annotation manuelle de corpus

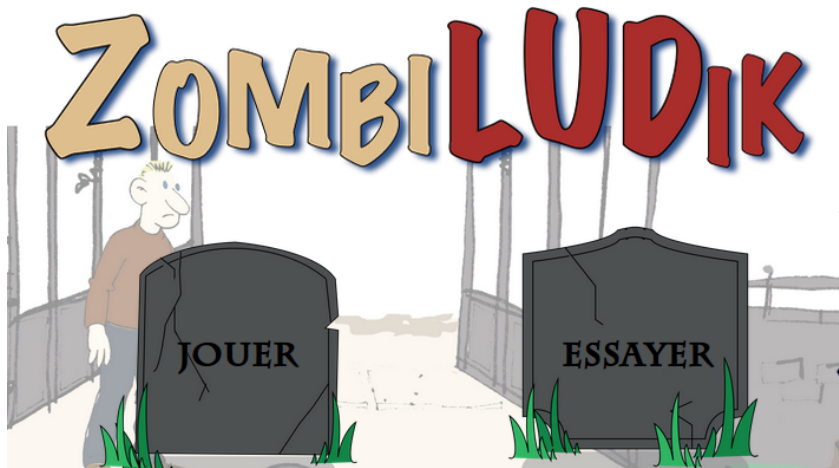
La production participative

ZombiLingo

**Conclusion**

Bientôt sur vos écran !

Le jeu pour Universal Dependencies : <http://zombiludik.org/>



# Ce que la production participative nous apprend

Il est possible d'outiller davantage l'annotation pour :

- ▶ former les annotateurs
- ▶ maintenir leur attention sur le temps long
- ▶ identifier et prendre en compte les (micro) spécialisations

mais également pour

- ▶ leur donner du *feedback* plus régulièrement
- ▶ leur fournir un espace d'échange (forum) entre eux et avec le gestionnaire de la campagne

THANK  YOU



<https://github.com/zombilingo>



<http://zombilingo.org/export>

# ZombiLingo : l'équipe et les financeurs



Bruno Guillaume  
(Sémagramme, Inria NGE)



Nicolas Lefèvre (Ingénieur)

*Inria*





## Appendix

Les jeux ayant un but

Motiver les différents types de joueurs

## Bibliographie

## Appendix

Les jeux ayant un but

Motiver les différents types de joueurs

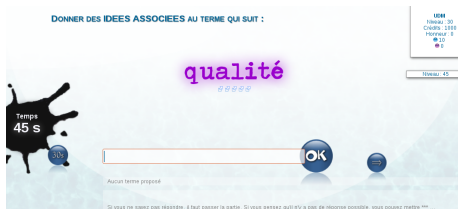
## Bibliographie

# JeuxDeMots : jouer à faire des associations d'idées. . .

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

Près de 300 million de relations (créées par env. 6 000 joueurs),  
mises à jour **constamment**

- ▶ jeu par pairs
- ▶ des relations de plus en plus complexes, typées
- ▶ des challenges
- ▶ des procès
- ▶ etc



# Phrase Detectives : jouer les détectives...

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

3,5 millions de décisions de  
45 000 joueurs (2016)

- ▶ corpus pré-annoté
- ▶ instructions détaillées
- ▶ formation
- ▶ 2 modes de jeu :
  - ▶ annotation
  - ▶ validation (correction d'annotations)

**DETECTIVES CONFERENCE**  
Another detective has made a decision about a phrase, either that 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?

**USERPROFILE**  
Ikaa  
23 who work  
1 decision  
21 agreements  
0 extras  
22 who month  
82 all time  
Level: Apprentice  
Your rating: 80%  
CASE OPEN  
32 tasks remaining  
1 completed case  
EDIT PROFILE / LOGOUT  
1 friend  
Ajouter le premier de vos amis à indiquer que vous

**SEARCH CLUES**  
Phrases that they like, like and that are likely to refer to the text so select more than one phrase if necessary.  
Words like they or them could refer to the text, one thing in the text so select more than one phrase if necessary.  
Always look for the closest previous mention of the phrase to score maximum agreement points.

**Knitta (Wikipedia)**  
PolyCoff and Miylk came up with their own names, then invented names for other members in a brainstorming session they considered "one of the more hilarious meetings". Some former member names include Krotosius N.T., SenOnStech and P-Krotk.  
As of January 2008, the group has two female members and one male, ages 30 to 73, who wish to remain anonymous. Current members' are PolyCoff, MascuOfinity, and Gravery SQ. An estimated five to twelve project groups exist around the world.  
Usually tagging on Friday nights and Sunday mornings, **knitta** taggers leave a paper tag on each work, bearing the slogan "kotta please" or "whaddup kotta?". They tag trees, lamp posts, railings, fire hydrants, monuments and other urban targets, and even get a little "hardcore" with ideas like hanging limited-edged breakers over aerial telephone cables. The crew marks holidays by doing themed work, using, for example, pink yarn for their Valentine's Day pieces and sparkly yarn for New Years. When Knitta is not working with a theme, they work on projects, tagging specific targets or specific areas.  
**the group** and their followers consider their graffiti "a method of beautifying public space".

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 this text or is it a property?  
Use your reason to select the **closest phrase(s)** if it has been mentioned before.

**USERPROFILE**  
Ikaa  
23 who work  
0 decisions  
21 agreements  
0 extras  
21 who month  
82 all time  
Level: Apprentice  
Your rating: 80%  
CASE OPEN  
37 tasks remaining  
1 completed case  
EDIT PROFILE / LOGOUT  
1 friend  
Ajouter le premier de vos amis à indiquer que vous

**SEARCH CLUES**  
Phrases beginning with a, an or the can refer to two different phrases.  
1. **an an object**  
They can be used to identify an object in the text, for example "The **professor**, delivered a letter" describes the object "Fred" as having the property of being "the professor".  
2. **is a property**  
They can also be used to say something about an object. For example "Fred, **the professor**, delivered a letter" describes the object "Fred" as having the property of being "the professor".  
If you think the phrase indicates a property try to select the closest phrase it refers to.

Not mentioned before This is a property

Done Feedback

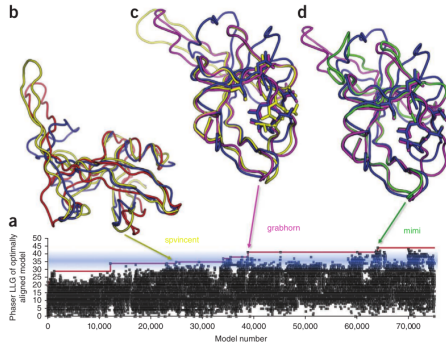
# FoldIt : jouer à replier des protéines. . .

. . . pour résoudre des problèmes de structure cristalline [Khatib et al., 2011]

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

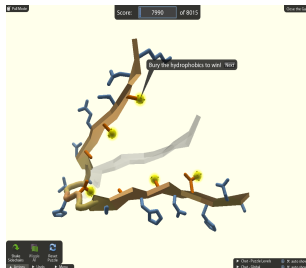
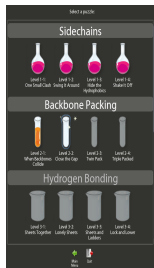
Solution à un problème non résolu depuis plus de 10 ans

- ▶ trouvée en quelques semaines
- ▶ par une équipe de joueurs
- ▶ qui va permettre la création de médicaments antirétroviraux



# FoldIt : jouer à replier des protéines. . .

. . . sans aucune connaissance préalable en biochimie [Cooper et al., 2010]



## Formation par étapes

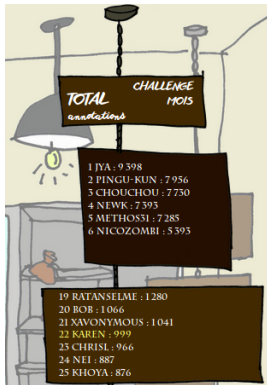
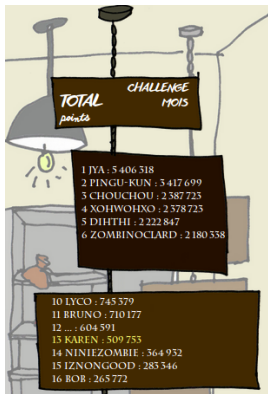
- ▶ tutoriel décomposé par concepts
- ▶ puzzles pour chaque concept
- ▶ l'accès aux puzzles suivants n'est octroyé que si le niveau du joueur est suffisant

# General features

Bring the fun through :

- ▶ zombie design
- ▶ use of (crazy) objects
- ▶ regular challenges (specific corpus and design) on a trendy topic :
  - ▶ Star Wars (when the movie was playing)
  - ▶ soccer (during the Euro)
  - ▶ Pokemon (well...)

# LeaderboardS (for achievers)

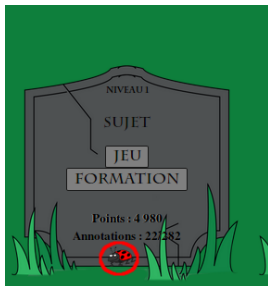


Criteria :

- ▶ number of annotations or points
- ▶ in total, during the month, during the challenge



## Hidden features (for explorers)



- ▶ appearing randomly
- ▶ with different effects : objects, other game, etc.

# Duels (for socializers (and killers?))

JOUER BOUTIQUE DUELS FORUM ADMIN JOUEURS

Duels Nouveau duel

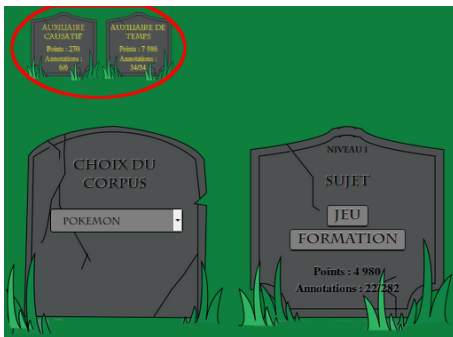
Duels gagnés : 1  
Duels perdus : 1  
Matches nuls : 1

DUELS LIBRES 1 MES DUELS EN COURS 2 MES DUELS TERMINÉS 3

	Karen vs. JYA 1420 1558	JYA gagne 5956 points	Complément en « de » duel en 20 en tours
	Karen vs. bruno 660 660	Karen gagne 660 points bruno gagne 660 points	Auxiliaire causatif duel en 20 en tours
	Karen vs. JYA 1754 1670	Karen gagne 3424 points	Complément en « de » duel en 20 en tours

- ▶ select an enemy
- ▶ challenge them on a specific type of relation

## Badges (?) (for collectors)



- ▶ play all the sentences for a relation type, for a corpus
- ▶ play all the sentences from a corpus



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

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

In Proceedings of the Traitement Automatique des Langues Naturelles (TALN), Grenoble, France.



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, Austria.



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.

-  Dandapat, S., Biswas, P., Choudhury, M., and Bali, K. (2009). Complex linguistic annotation - no easy way out ! a case from bangla and hindi POS labeling tasks.  
In Proceedings of the third ACL Linguistic Annotation Workshop, Singapore.
-  Deneme (2009).  
How many turkers are there ?  
<http://groups.csail.mit.edu/uid/deneme/>.
-  Desrosières, A. (1989).  
Comment faire des choses qui tiennent : histoire sociale et statistique.  
Histoire & Mesure, 4(4) :225–242.
-  Desrosières, A. (2001).  
Entre réalisme métrologique et conventions d'équivalence : les ambiguïtés de la sociologie quantitative.  
Genèses, 2(43) :112–127.



Desrosières, A. (2008).

Pour une sociologie historique de la quantification :  
L'Argument statistique I.

Presses de l'École des Mines de Paris.



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., Nazarenko, A., and Rosset, S. (2012).

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

In International Conference on Computational Linguistics  
(COLING), pages 895–910, Mumbai, India.



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

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

In Fourth ACL Linguistic Annotation Workshop, pages 56–63,  
Uppsala, Sweden.



Guillaume, B. and Perrier, G. (2015).  
Dependency Parsing with Graph Rewriting.

In

Proceedings of IWPT 2015, 14th International Conference on Parsing  
pages 30–39, Bilbao, Spain.



Habert, B. (2005).  
Portrait de linguiste(s) à l'instrument.  
Texto!, vol. X(4).



Howe, J. (2006).  
The rise of crowdsourcing.  
Wired Magazine, 14(6).



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.



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

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

In Proceedings of the Journées internationales d'Analyse statistique des Données Textuelles (JADT), Lyon, France.



Leech, G. (1997).

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



Nédellec, C., Bessières, P., Bossy, R., Kotoujansky, A., and Manine, A.-P. (2006).

Annotation guidelines for machine learning-based named entity recognition in microbiology.

In et C. Nédellec, M. H., editor, Proceedings of the Data and text mining in integrative biology workshop, pages 40–54,  
Berlin, Germany.



Urieli, A. (2013).



Robust French syntax analysis : reconciling statistical methods  
and linguistic knowledge in the Talismane toolkit.

PhD thesis, Université de Toulouse II le Mirail, France.