



Crowdsourcing Complex Language Resources: Playing to Annotate Dependency Syntax

Bruno Guillaume, **Karën Fort** and Nicolas Lefèbvre

bruno.guillaume@inria.fr, karen.fort@paris-sorbonne.fr, nicolas.lefebvre@inria.fr

December 16th, 2016



- 1 Crowdsourcing: back to basics
- 2 Overview of the game
- 3 Behind the curtain
- 4 Results
- 5 Conclusion and future plans

- 1 Crowdsourcing: back to basics
- 2 Overview of the game
- 3 Behind the curtain
- 4 Results
- 5 Conclusion and future plans

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]

- no *a priori* identification or selection of the participants ("open call")

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]

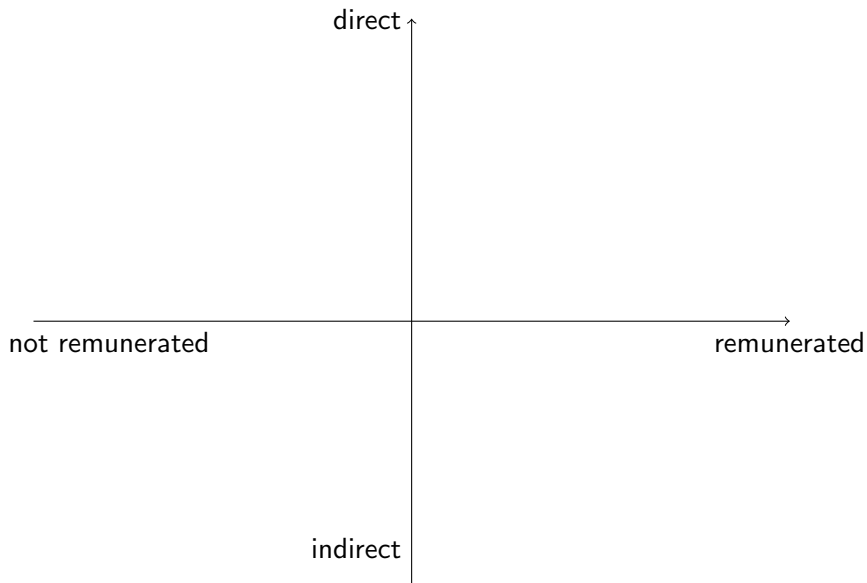
- **no *a priori* identification** or selection of the participants ("open call")
- **massive** (in production and participation)

Crowdsourcing

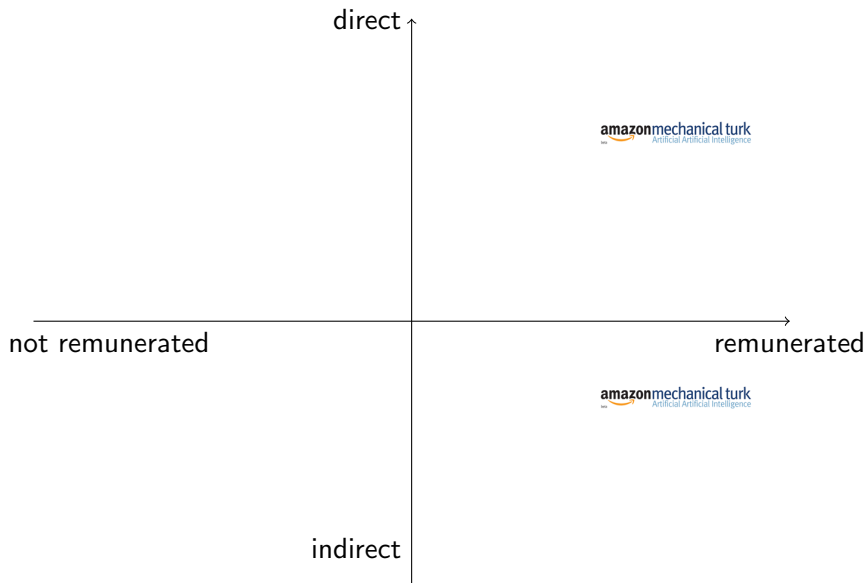
Crowdsourcing is *"the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."* [Howe, 2006]

- no *a priori* identification or selection of the participants ("open call")
- massive (in production and participation)
- (relatively) cheap

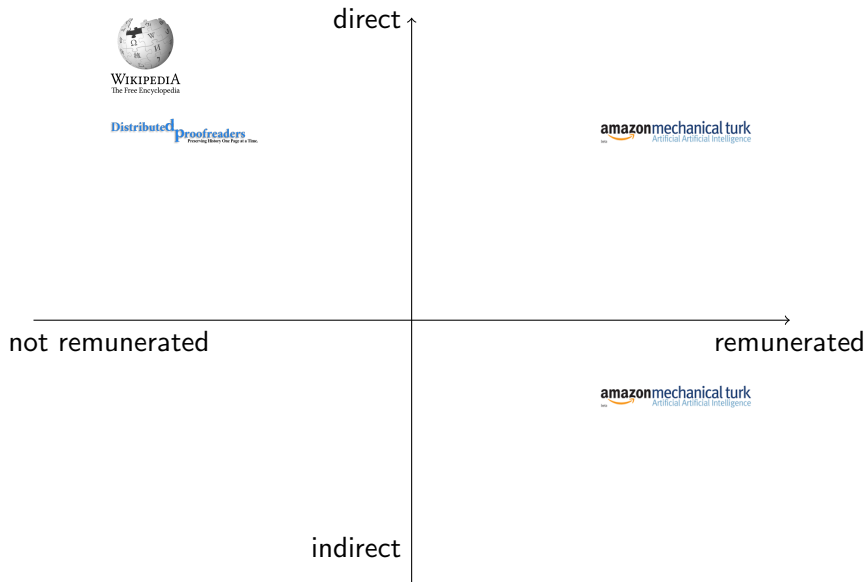
A simplified taxonomy (more in [Geiger et al., 2011])



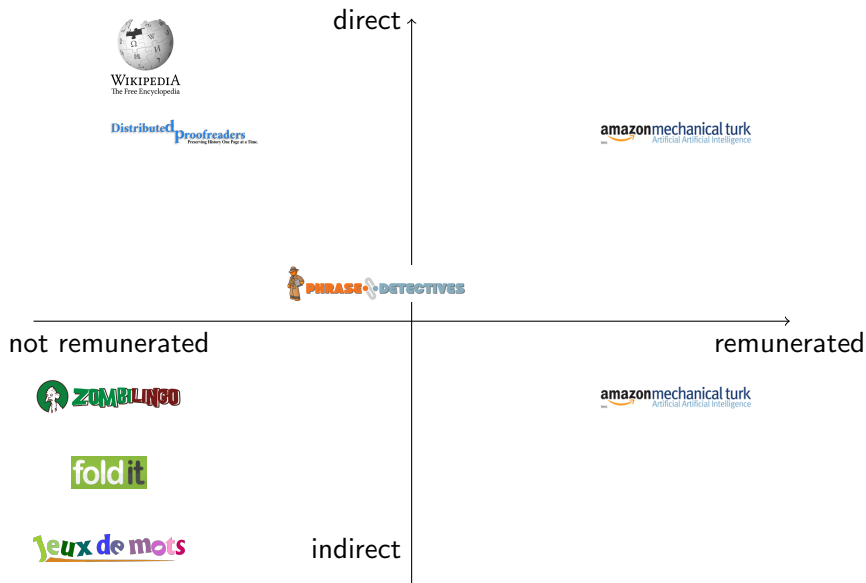
A simplified taxonomy (more in [Geiger et al., 2011])



A simplified taxonomy (more in [Geiger et al., 2011])



A simplified taxonomy (more in [Geiger et al., 2011])

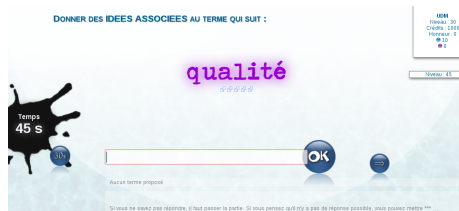


JeuxDeMots: playing association of ideas...

...to create a lexical network [Lafourcade and Joubert, 2008]

More than **60 million relations** (created by 1,161 players), that are **constantly** updated

- play by pairs
- more and more complex, typed relations
- challenges
- lawsuits
- etc.



Phrase Detectives: playing detective...

...to annotate co-reference [Chamberlain et al., 2008]

3.5M decisions from 45k players

- pre-annotated corpus
- detailed instructions
- training
- 2 different playing modes
 - ▶ annotation
 - ▶ validation (correction of 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?

Knittia (Wikipedia)

PolyCoti and Akyke 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 Kneious N.I.T., SonOfDitch and P-Knity.

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 PolyCoti, MascuKnitty, and Granny SQ. An estimated five to twelve copycat groups exist around the world.

Usually tagging on Friday nights and Sunday mornings, **Knitta tagged** leave a paper tag on each work, bearing the slogan "knitta please" or "whaddup knitta?" They tag trees, lamp posts, railings, fire hydrants, monuments and other urban targets, and even get a little "hardcore" with ideas like hanging knitted-bagged smokers over aerial telephone cable. The crew marks holidays by doing themed work, using, for example, pink yarn for their Valentine's Day pieces and spindly 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.

NAME THE CULPRIT

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

Knittia (Wikipedia)

PolyCoti and Akyke 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 Kneious N.I.T., SonOfDitch and P-Knity.

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 PolyCoti, MascuKnitty, and Granny SQ. An estimated five to twelve copycat 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 "knitta please" or "whaddup knitta?" They tag trees, lamp posts, railings, fire hydrants, monuments and other urban targets, and even get a little "hardcore" with ideas like hanging knitted-bagged smokers over aerial telephone cable. The crew marks holidays by doing themed work, using, for example, pink yarn for their Valentine's Day pieces and spindly yarn for New Years. When Knitta is not working with a theme, they work on projects, tagging specific targets or specific areas.

Not mentioned before ☐ **This is a property**

SEARCH CLUES

Words like they, her and it are ways to refer to something else in the text. Try to find the closest mention of the phrase.

Words like they or them could refer to more than one thing in the text so select more than one phrase if necessary.

Always look for the **closest phrase** number of the phrase to score maximum agreement points.

1. As an object
This can be used to identify an object in the text, for example "The postman delivered a letter" describes the object "letter" in having the property of being "The postman".

2. As a property
They can also be used to say something about an object. For example "The postman, delivered a letter" describes the object "The postman" in having the property of being "The postman".

If you think the phrase describes a property try to select the closest phrase it refers to.

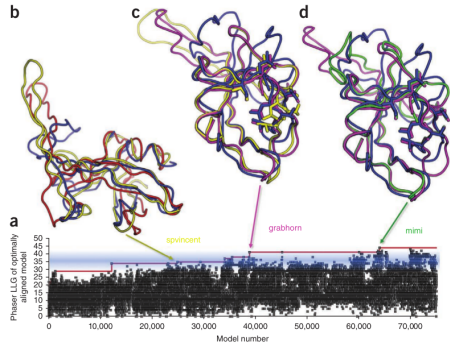
FoldIt: playing proteins folding...

...to solve scientific issues [Khatib et al., 2011]

Solution to the crystal structure of a monomeric retroviral protease
(simian AIDS-causing monkey virus)

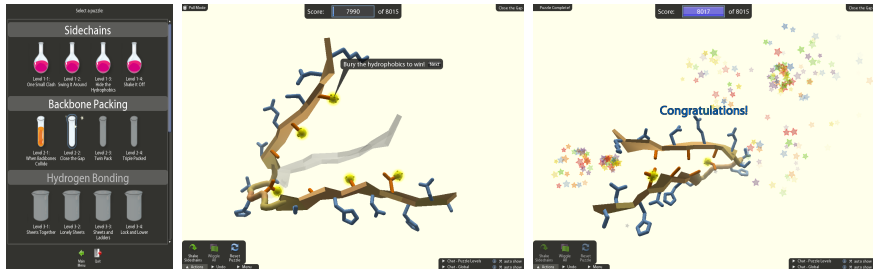
Solution to an issue unsolved for over
a decade

- found in a couple of weeks
- by a team of players
- that will allow for the creation of antiretroviral drugs



FoldIt: playing proteins folding...

...without any prior knowledge in biochemistry [Cooper et al., 2010]

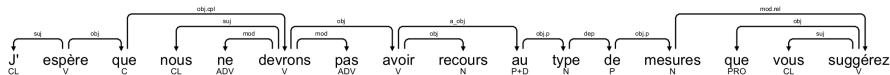


Step-by-step training

- tutorial decomposed by concepts
- puzzles for each concept
- access to the following puzzles is given only if your level is sufficient

- 1 Crowdsourcing: back to basics
- 2 Overview of the game**
- 3 Behind the curtain
- 4 Results
- 5 Conclusion and future plans

A complex annotation task



- annotation guidelines

- ▶ 29 relation types
- ▶ approx. 50 pages

- counter-intuitive decisions: aobj = au

[...] avoir recours au type de mesures [...]

i.e. head of the PP is the preposition

→ **decompose** the complexity of the task [Fort et al., 2012],
not simplify it!

<http://zombilingo.org/>

ZOMBI LINGO

BIENVENUE à TOI, JEUNE ZOMBIE !

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

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

ATTENTION AUX
PIÈGES, ILS SONT
NOMBREUX !



Jouer

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

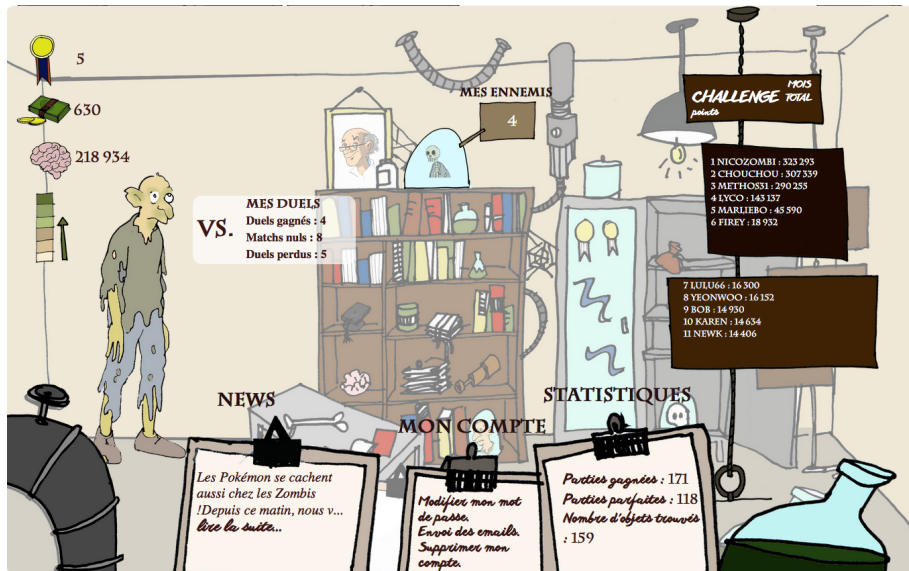
Karen

*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





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.

74

43

57
 150
 Acheter

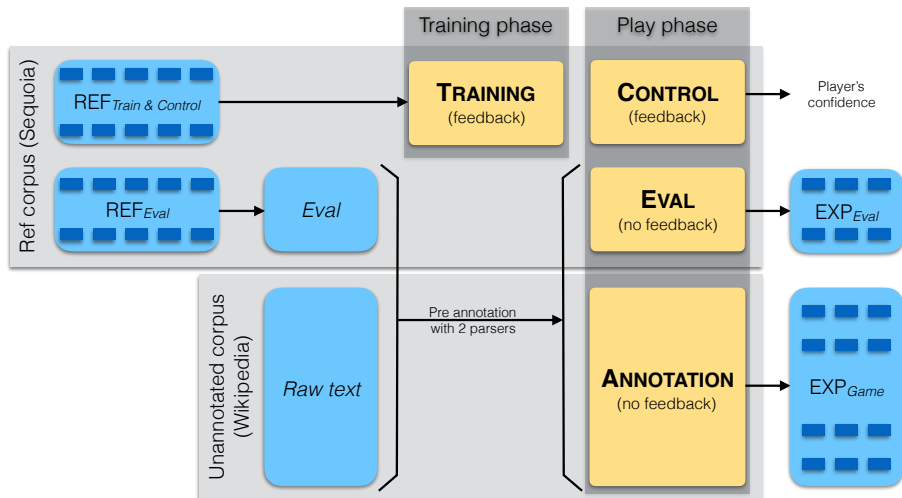
1
 15
 Acheter

3
 300
 Acheter

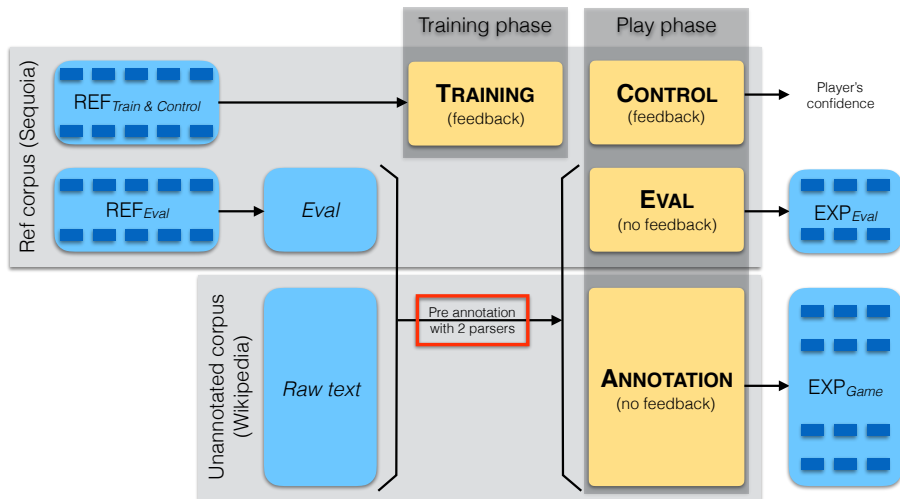
0
 15
 Acheter

- 1 Crowdsourcing: back to basics
- 2 Overview of the game
- 3 Behind the curtain**
- 4 Results
- 5 Conclusion and future plans

Organizing quality assurance



Preprocessing data (freely available corpora)



Preprocessing data (freely available corpora)

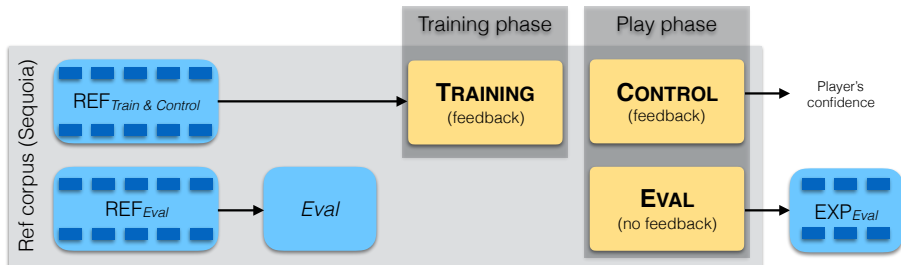
Pre-annotation with two parsers

- ➊ a statistical parser: Talismane [Urieli, 2013]
- ➋ a symbolic parser, based on graph rewriting:
FRDEP-PARSE [Guillaume and Perrier, 2015]

→ play the items for which the two parsers give different annotations

Training, control and evaluation

Reference: 3,099 sentences of the Sequoia corpus [Candito and Seddah, 2012]



$REF_{Train \& Control}$	REF_{Eval}	Unused
50%	25%	25%
1,549 sentences	776 sentences	774 sentences

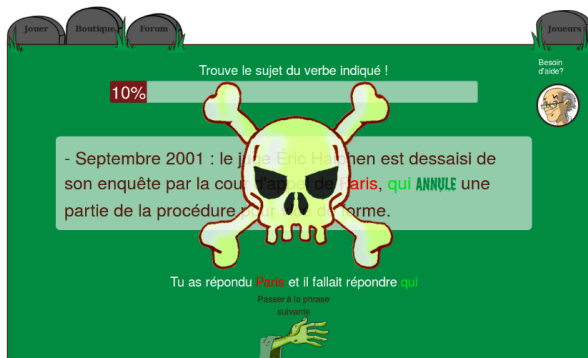
- $REF_{Train \& Control}$ is used to train the players
- REF_{Eval} is used like a raw corpus, to evaluate the produced annotations

Training the players

Compulsory for each dependency relation



- sentences are taken from the REF_{Train&Control} corpus
- a feedback is given in case of error



Dealing with cognitive fatigue and long-term players

Control mechanism

Sentences from the REF_{Train&Control} corpus are proposed regularly

- 1 if the player fails to find the right answer, a feedback with the solution is given

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



Dealing with cognitive fatigue and long-term players

Control mechanism

Sentences from the REF_{Train&Control} corpus are proposed regularly

- 1 if the player fails to find the right answer, a feedback with the solution is given
- 2 after a given number of failures on the same relation, the player cannot play anymore and has to redo the corresponding training

- 1^{er} **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](#)



Dealing with cognitive fatigue and long-term players

Control mechanism

Sentences from the $\text{REF}_{\text{Train\&Control}}$ corpus are proposed regularly

- ➊ if the player fails to find the right answer, a feedback with the solution is given
 - ➋ after a given number of failures on the same relation, the player cannot play anymore and has to redo the corresponding training
- we deduce a **level of confidence** for the player on this relation

- 1 Crowdsourcing: back to basics
- 2 Overview of the game
- 3 Behind the curtain
- 4 Results**
- 5 Conclusion and future plans

Production: game corpus size

compared to other existing French dependency syntax corpora

As of July 10, 2016

- 647 players
- who produced 107,719 annotations

	Sequoia 7.0	UD-French 1.3	FTB-UC	FTB-SPMRL	Game
Sentences	3,099	16,448	12,351	18,535	5,221
Tokens	67,038	401,960	350,947	557,149	128,046
Tokens/sent.	21.6	24.4	28.4	30.1	24.5

Production: game corpus size

compared to other existing French dependency syntax corpora

As of July 10, 2016

- 647 players
- who produced 107,719 annotations

	Sequoia 7.0 <i>free</i>	UD-French 1.3 <i>free</i>	FTB-UC <i>not "free"</i> ¹	FTB-SPMRL <i>not "free"</i>	Game <i>free</i>
Sentences	3,099	16,448	12,351	18,535	5,221
Tokens	67,038	401,960	350,947	557,149	128,046
Tokens/sent.	21.6	24.4	28.4	30.1	24.5

¹No redistribution allowed.

Production: game corpus size

compared to other existing French dependency syntax corpora

As of July 10, 2016

- 647 players (775 as of Dec. 13th)
- who produced 107,719 annotations (168,832 as of Dec. 13th)

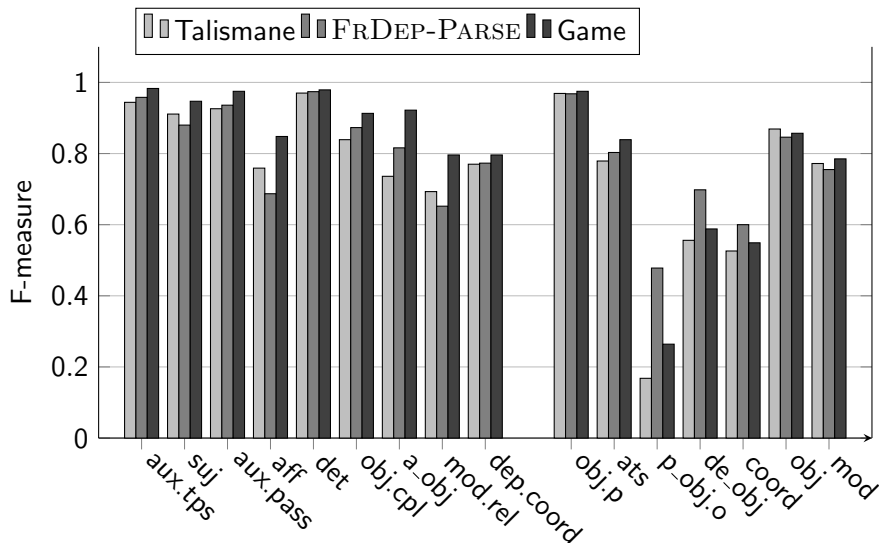
	Sequoia 7.0 <i>free validated</i>	UD-French 1.3 <i>free after ZL¹ + errors</i>	FTB-UC <i>not "free" validated</i>	FTB-SPMRL <i>not "free" validated</i>	Game <i>free validated</i>
Sent.	3,099	16,448	12,351	18,535	5,221
Tok.	67,038	401,960	350,947	557,149	128,046
Tok./sent.	21.6	24.4	28.4	30.1	24.5

+ (ever)growing resource!

¹ZL 1.0, July 2014 vs UD 1.0 January 2015.

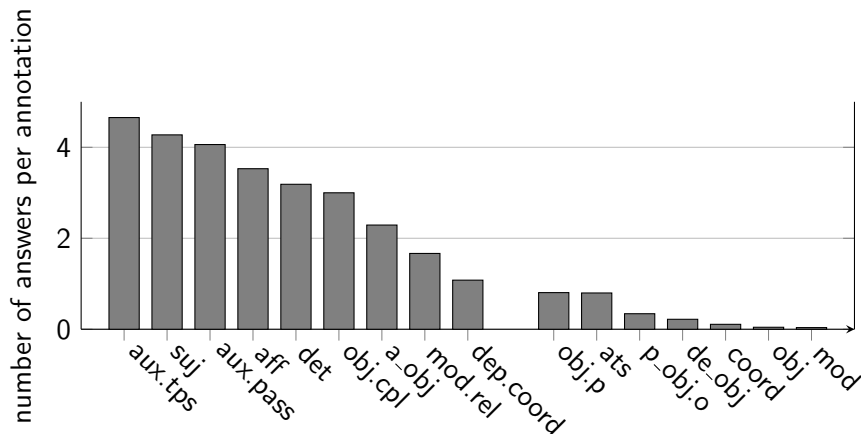
Evaluating quality

on the REF_{Eval} corpus



Annotation density

on the REF_{Eval} corpus



→ need **more** annotations on some relations

- 1 Crowdsourcing: back to basics
- 2 Overview of the game
- 3 Behind the curtain
- 4 Results
- 5 Conclusion and future plans**

Games With(out) A Problem [Tuite, 2014]?

Achievements

- surprisingly good results in terms of quantity and **quality**
- we demonstrated that **we can** train people on a complex task

Yet to be validated

- relation types which are not played (too difficult or lack of players?)

Difficulties

- communication / advertisement
- community management

Improving gamification

- give more to explore and collect
- build a real story
- build a sense of community (how?)

Improving the exported resource

Test the influence of

- the pre-annotation score
- the level of the player in the game
- the confidence we have in the player for the relation type at hand

Expand to new languages and new annotation types

New languages

- English
- **less-resourced languages**

Alice Millour (PhD student)



New annotation types

- **part-of-speech** (POS),
- corpus building,
- etc.

Team and fundings



Bruno Guillaume (researcher)



Nicolas Lefebvre (engineer)



THANK  YOU



<https://github.com/zombilingo>



<http://zombilingo.org/export>



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

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

In Proceedings of 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).



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.



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

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

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



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.



Guillaume, B. and Perrier, G. (2015).

Dependency Parsing with Graph Rewriting.

In

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



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



Tuite, K. (2014).

Gwaps: Games with a problem.

Foundations of Digital Games 2014.



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.