



CRÉATION DE RESSOURCES LINGUISTIQUES PAR APPROCHES COOPÉRATIVES

JEUX CONSTRUCTIFS ET APPROCHES COOPÉRATIVES NON LUDIQUES

Fabrizio Gotti – gottif@iro.umontreal.ca

Philippe Langlais – felipe@iro.umontreal.ca

15 JUILLET 2013

RALI

Recherche Appliquée
en Linguistique Informatique
rali.iro.umontreal.ca

Université 
de Montréal

Sommaire

Plusieurs tâches simples pour les êtres humains confondent encore les ordinateurs. En traitement des langues naturelles (TALN), ces défis incluent la reconnaissance de la parole, la traduction, la résolution d'ambiguïtés de sens, etc. Pour concevoir des programmes exécutant ces tâches automatiquement, il faut le plus souvent les *entraîner* sur des données conçues par des êtres humains. Traditionnellement, des experts les élaborent, à des coûts qui deviennent de plus en plus prohibitifs, de l'ordre de 1 M USD pour 1 M de mots annotés.

Avec l'avènement de la coopération rendue possible *via* Internet, il est désormais possible de tabler sur une large communauté d'utilisateurs en ligne pour répartir ce travail de construction. On distingue trois stratégies différentes de coopération.

1. Les **approches volontaires bénévoles** bénéficient de l'altruisme de milliers d'utilisateurs qui, comme pour Wikipédia, élaborent des ressources au nom des *sciences citoyennes*. Le wiktionnaire est un exemple de ce genre de produit de très haute qualité.
2. Le **jeu constructif (*game with a purpose*)**, plus récent, propose au participant un jeu vidéo engageant, dont un des sous-produits est la création de ressources linguistiques. Ces jeux canalisent ainsi une fraction des centaines de millions d'heures jouées par jour dans le monde. Par exemple, le jeu Phrase Detectives a permis à 3000 joueurs d'abattre 5000 heures de travail bénévole, pour colliger 1,3 M d'éléments d'information.
3. Des **plates-formes de *crowdsourcing*** comme Amazon Mechanical Turk permettent à un maître d'œuvre de segmenter un travail en plusieurs microtâches soumises à des utilisateurs, contre une très faible rémunération (quelques sous par microtâche).

Ces approches ont en commun la nécessité de concevoir l'interface de collaboration en tenant compte de la motivation des utilisateurs, et en encourageant la qualité, souvent en recourant à la validation entre utilisateurs, et à divers artifices d'interface graphique.

Les approches volontaires bénévoles et les jeux constructifs doivent de plus inciter une participation conséquente. Les jeux constructifs utilisent notamment les incitatifs classiques des jeux vidéo, soit l'amour du jeu (attisé par divers procédés ludiques enthousiasmant les joueurs), le réseau social (facilité par des plates-formes comme Facebook qui encouragent la compétition amicale et la coopération), et de petites récompenses en argent.

Les études ayant abordé la compétitivité financière de ces approches coopératives indiquent bien que celles-ci permettent des économies considérables par rapport aux approches traditionnelles. Les stratégies bénévoles sont d'un grand intérêt dans l'élaboration de ressources linguistiques de grande taille et de qualité, alors que les plates-formes avec rémunération sont très rentables pour les tâches plus restreintes. La qualité des ressources obtenues rivalise avec celle des approches traditionnelles.

Table des matières

Sommaire.....	ii
Table des matières.....	iii
Liste des illustrations.....	v
Liste des tableaux.....	vi
1 Introduction.....	1
2 Les ressources linguistiques en traitement des langues naturelles.....	2
2.1 Le traitement automatique des langues naturelles (TALN).....	2
2.2 Les ressources linguistiques.....	2
2.3 Élaboration traditionnelle de ressources linguistiques.....	3
3 Brève typologie des approches coopératives.....	5
3.1 Contributions involontaires.....	5
3.2 Contributions volontaires bénévoles.....	6
3.3 Contributions volontaires rémunérées.....	6
3.4 Contributions par le jeu constructif.....	7
4 Quelques distinctions : ludification, jeu sérieux et jeu constructif.....	8
5 Deux exemples de jeux constructifs.....	11
5.1 JeuxDeMots.....	11
5.2 Phrase Detectives.....	13
6 Conception des jeux constructifs.....	15
6.1 Trois mécaniques de fonctionnement.....	15
6.1.1 Jeu avec accord sur les sorties (<i>output-agreement games</i>).....	15
6.1.2 Jeu avec inversion du problème (<i>inversion-problem games</i>).....	15
6.1.3 Jeu avec accord sur les entrées (<i>input-agreement games</i>).....	16
6.1.4 Variantes possibles.....	16
6.1.5 Partenaire de jeu artificiel (« bot »).....	17
6.2 Interaction avec le joueur.....	17
6.3 Incitatifs au jeu constructif.....	17
6.3.1 Quelques théories psychologiques.....	17
6.3.2 Stratégies incitatives favorisant le plaisir du jeu.....	18

6.3.3	Stratégies incitatives liées aux aspects sociaux du jeu.....	21
6.3.4	Incitatifs financiers : un pari risqué ?.....	22
6.4	Jeu constructif et médias sociaux	23
7	Qualité des annotations.....	25
7.1	Métriques de qualité	25
7.2	Agrégation des annotations et validation	26
7.3	Principes de conception pour des annotations de qualité	27
7.4	Principes de conception contre la collusion et les abus.....	30
8	Approches coopératives non ludiques	32
8.1	Motivation des utilisateurs.....	32
8.2	Exemples d'approches coopératives non ludiques bénévoles	32
8.3	Exemple d'approches coopératives non ludiques rémunérées	35
8.4	Comparaison avec le jeu constructif.....	36
8.4.1	Différences.....	36
8.4.2	Similarités.....	37
9	Compétitivité financière des approches coopératives	38
9.1	Dépenses en approches coopératives.....	38
9.2	Études comparatives [5], [9]	39
9.3	Étude de Thaler et coll. [30].....	40
9.4	Conclusions.....	41
10	Conclusion et perspectives.....	42
11	Références.....	44
	Annexe A : Quelques jeux constructifs en bref	47
	Annexe B : Quelques jeux constructifs non linguistiques	48
	Annexe C : Quelques approches coopératives non ludiques	50

Liste des illustrations

Figure 1 – Extrait des annotations du corpus Brown, pour la phrase « At the same time reaction among anti-organization Democratic leaders and in the Liberal party to the Mayor’s reported plan was generally favorable. ». Chaque mot est accompagné d’un code indiquant sa partie du discours.	3
Figure 2 – Un exemple de reCaptcha lors de l’inscription sur un site web.....	5
Figure 3 – Interface de classification de cratères lunaires sur Zooniverse.org	6
Figure 4 – Jeu ESP Game	7
Figure 5 – Exemple de question posée sur www.stackoverflow.com montrant les points et badges des utilisateurs.....	8
Figure 6 – The Lottery of Life simule la naissance de l’internaute dans un pays du monde choisi au hasard et lui explique les difficultés des enfants dans « son » pays.	9
Figure 7 – Dans les jeux Zombie Typocalypse (haut) et Word Shoot (bas), le joueur repousse des hordes d’ennemis en dactylographiant rapidement les étiquettes de ces attaquants afin de les abattre.	10
Figure 8 – Interface de jeu de JeuxDeMots, pour des idées associées au terme « reportage »	11
Figure 9 – Extrait du réseau lexico-sémantique construit par JeuxDeMots.....	12
Figure 10 – Partie de Phrase Detectives, pour l’article Wikipédia <i>Les Aventures d’Alice au pays des merveilles</i> , en mode « Identifiez le coupable »	14
Figure 11 – Partie de Phrase Detectives, en mode de validation.....	14
Figure 12 – Partie du jeu constructif Verbosity, tiré de [12].....	16
Figure 13 – Interface de jeu GIVE.....	19
Figure 14 – Exemple de rétroaction positive pour le jeu Wordrobe.....	19
Figure 15 – Progression du niveau « recrue » à « détective » dans Phrase Detectives	20
Figure 16 – Partie du jeu de société Scattergories™ pour la lettre « B » (haut) et interface de tutoriel de Categorilla (bas)	21
Figure 17 – Palmarès des meilleurs joueurs dans Wordrobe	22
Figure 18 – Interface du jeu Sentiment Quiz, tirée de [22]	24
Figure 19 – Jeu AKI (http://www.jeuxdemots.org/AKI.php)	26
Figure 20 – Jeu constructif 1001 Paraphrases pour la reformulation de « this can help you »	28
Figure 21 – Interface du jeu <i>Senses</i> de la famille de jeux constructifs Wordrobe.....	30
Figure 22 – Tricherie dans le jeu Verbosity commise par le partenaire (panneau de droite)	31
Figure 23 – Extrait de l’entrée vietnamienne du mot « papillon » dans le projet Papillon.....	33
Figure 24 – Traductions vers le français de Facebook, dans Facebook.....	34
Figure 25 – « Arbre de compétences » pour l’apprentissage du français sur Duolingo	35
Figure 26 – Le jeu OntoPronto. Capture d’écran issue de la vidéo YouTube.	40

Liste des tableaux

Tableau 1 – Exemples de relations lexicales recueillies par JeuxDeMots, tiré de [16]	12
Tableau 2 – Accord sur les annotations de Phrase Detectives, par type, tiré de [5]	25
Tableau 3 – Coûts d’annotation, pour Phrase Detectives et JeuxDeMots, tiré de [9]. Le nombre de mois indique le nombre de mois nécessaire au développement + le nombre de mois de fonctionnement.	39
Tableau 4 – OntoPronto versus Mechanical Turk, tiré de [30]	41
Tableau 5 – Tableau synthétique de quelques jeux constructifs. Les indications entre crochets après les noms de jeux sont respectivement la référence bibliographique, une figure dans ce texte et une url vers le jeu.	47
Tableau 6 – Quelques approches coopératives non ludiques en création de ressources linguistiques.....	50

1 Introduction

En dépit des progrès remarquables faits dans le domaine de l'intelligence artificielle au cours des dernières décennies, l'informatique cherche encore des solutions à des problèmes qui s'avèrent très simples à résoudre pour les êtres humains. En traitement des langues, le champ de recherche qui nous intéresse ici, les illustrations abondent. Ainsi, dans la phrase « L'avocat est délicieux dans les salades. », il est évident pour un humain que l'on parle du fruit « avocat » et non de la profession. Pourtant, ce problème de désambiguïsation sémantique (c'est-à-dire de sens des mots) est complexe pour un programme informatique.

Actuellement, pour faire apprendre à un logiciel ce genre de subtilités, on recourt souvent à une phase d'apprentissage, qui consiste à exposer le programme à des informations pertinentes. Ces ressources sont souvent volumineuses. On estime par exemple qu'un adulte moyen possède plusieurs millions d'informations simples qui constituent sa « connaissance du monde » [1]. Quand elle est envisageable, la constitution de ces ressources est par conséquent un effort onéreux et de longue haleine. Pourtant, les applications importantes ne manquent pas : traduction automatique (où l'on traduira notre exemple par « avocado » plutôt que par « lawyer »), analyse syntaxique, moteurs de recherche, etc.

Les ressources linguistiques sont traditionnellement élaborées par des experts. Cependant, l'avènement d'Internet voit apparaître des stratégies nouvelles de coopération, par lesquelles des internautes réunis autour d'une cause commune ont su édifier collectivement des ressources d'une grande qualité. Wikipédia en est peut-être l'exemple le plus frappant, avec ses 1,4 million d'articles en français, dont la qualité rivalise avec les encyclopédies plus traditionnelles, selon une étude de *Nature* [2]. Le wiktionnaire^a, projet satellite de dictionnaire, est déjà utilisé à profit en traitement des langues.

En 2004, le chercheur Luis von Anh propose de canaliser les forces des foules à l'aide d'un jeu vidéo. Son jeu ESP [3] propose à deux joueurs d'essayer de s'entendre sur des mots décrivant une image, avec, comme sous-produit désirable, l'étiquetage bénévole de 300 000 images avec des mots-clés utiles aux moteurs de recherche. Depuis lors, ces *jeux constructifs* (*games with a purpose*, ou GWAP) se multiplient, et profitent d'une fraction de la manne des millions d'heures-personnes investies dans le jeu vidéo annuellement.

Dans ce travail, nous proposons de décrire les approches coopératives de construction de ressources linguistiques, en nous concentrant sur les jeux constructifs. Après avoir brièvement décrit les ressources linguistiques en traitement des langues (section 1), nous définirons et situerons les différentes approches coopératives actuelles, en sections 3 et 4. Aux sections 5, 6 et 7, nous expliquerons les principes de fonctionnement des jeux constructifs, les ressources qu'ils permettent de créer, et leurs limitations. En section 8, nous esquisserons le portrait des approches coopératives non ludiques, avant de conclure sur les mérites financiers de chaque stratégie, à la section 9.

^a <https://fr.wiktionary.org/wiki/>

2 Les ressources linguistiques en traitement des langues naturelles

2.1 Le traitement automatique des langues naturelles (TALN)

Dans ce travail, nous nous intéressons aux ressources utilisées dans le champ de recherche scientifique qu'est le traitement automatique des langues naturelles (TALN). Les experts du domaine se penchent sur divers problèmes au confluent de l'informatique et de la linguistique, dont l'importance économique n'est plus à démontrer.

On reconnaît ainsi les champs d'application suivants :

- Extraction d'information (recherche d'informations-clés) dans un document ;
- Moteurs de recherche ;
- Correction grammaticale (comme le correcteur de Microsoft Word, par exemple) ;
- Traduction automatique, et outils d'aide à la traduction ;
- Résumé automatique ;
- Reconnaissance de la parole ;
- Génération de texte ;
- Etc.

La résolution de ces tâches a largement profité de la révolution statistique qui s'est produite au sein de la communauté de TALN depuis quelques décennies. Ces approches ont cependant créé un besoin important de données linguistiques annotées, afin d'*entraîner* et d'évaluer les systèmes informatiques ainsi conçus. Ces systèmes doivent en effet être entraînés pour apprendre des humains comment effectuer leur tâche. Par exemple, un système voué à déterminer les parties du discours d'un texte (nom, verbe, préposition, etc.) pourra profiter d'un ensemble de textes étiquetés par un expert humain.

2.2 Les ressources linguistiques

Le travail de création des ressources linguistiques prend la forme d'*annotations* liées au texte faisant l'objet de l'étude, que l'on appelle *corpus*. Ces ressources sont à ce point précieuses qu'une conférence (LREC^b) est tenue tous les deux ans sur le sujet.

Ces annotations sont aussi diverses que les champs du TALN, évoqués à la section précédente. On y retrouve notamment :

- Annotations d'**entités nommées** (par exemple, juges, parties, verdicts, dans un corpus juridique) ;
- Annotations de **résolution d'anaphore**, où sont indiqués les antécédents des pronoms ambigus et de certaines entités textuelles ;
- **Traductions** humaines de textes, ou encore révisions manuelles de traductions ;
- Annotations sur la **qualité de résumés** ;

^b <http://www.lrec-conf.org/>

- **Réseaux lexicaux**, où l'on relie différents mots et termes par les relations sémantiques qui les unissent (par exemple, antonymie, synonymie, « est constitué de », etc.) ;
- **Ontologies**, où l'on tente d'établir manuellement une hiérarchie de concepts (par exemple, « un chat est un type de félin »), etc.
- Phrases annotées avec le **sentiment** (positif ou négatif) qui s'en dégage ;
- Axiomes de la **connaissance du monde** (*common sense knowledge*), déjà invoqués en introduction ;
- Etc.

2.3 Élaboration traditionnelle de ressources linguistiques

L'approche traditionnelle de création de ces ressources fait appel à un ou plusieurs experts du domaine afin de produire « manuellement » les annotations nécessaires.

Ce genre de projet requiert que l'on s'entende sur une méthodologie d'annotation, que les annotateurs soient correctement formés et, parfois, qu'un processus d'arbitrage soit possible.

Pour en assurer la qualité, ces annotations peuvent être *validées* entièrement, c'est-à-dire révisées par un (autre) expert. Lorsque cet effort est jugé prohibitif, certaines ressources sont échantillonnées et ces sous-ensembles sont validés pour détecter d'éventuels problèmes systématiques, tels que l'incompréhension des instructions d'annotation, qui auraient pu entacher la ressource significativement.

Un des premiers corpus annotés a été le corpus Brown [4], comptant 1 M de mots dont la partie du discours (verbe, article, nom, etc.) et le lemme (la forme primitive d'un mot) sont indiqués dans les annotations (voir Figure 1). Ce travail manuel a encore cours aujourd'hui, en particulier pour l'annotation de données linguistiques plus complexes [5].

```
At/in the/at same/ap time/nn reaction/nn among/in anti-organization/jj Democratic/jj-
tl leaders/nns and/cc in/in the/at Liberal/jj-tl party/nn to/in the/at Mayor's/nn$-tl
reported/vbn plan/nn was/bedz generally/rb favorable/jj ./.
```

Figure 1 - Extrait des annotations du corpus Brown, pour la phrase « At the same time reaction among anti-organization Democratic leaders and in the Liberal party to the Mayor's reported plan was generally favorable. ». Chaque mot est accompagné d'un code indiquant sa partie du discours.

Les coûts sont élevés : de l'ordre de 1 M USD pour un corpus d'un million de mots [5]. Et bien qu'il existe encore des bailleurs de fonds prêts à investir ces sommes en annotation manuelle (par exemple les projets Salsa^c et OntoNotes [6]), certaines études montrent qu'un million de mots est insuffisant pour entraîner des systèmes sur certaines tâches complexes et essentielles comme l'analyse syntaxique (projet du Penn Treebank). Les langues naturelles permettent en effet d'innombrables variations et ambiguïtés, et un million de

^c <http://www.coli.uni-saarland.de/projects/salsa/page.php?id=index>

mots ne suffit simplement pas à *couvrir* tous les cas possibles pour en informer les algorithmes informatiques.

Ces obstacles financiers limitent la construction de ressources très utiles. Par exemple, une étude révèle [7] que le projet Electronic Dictionary Research (EDR) visant l'élaboration d'un dictionnaire japonais-anglais a nécessité plus de 1200 années-personnes de travail. Son prix de vente (environ 19 000 CAD) est très inférieur aux coûts réels de sa construction, ce qui condamne l'aventure à un déficit ruineux. Les acheteurs potentiels sont également limités : trop cher pour un particulier, seules quelques institutions spécialisées peuvent acquérir le dictionnaire.

Qui plus est, pour atteindre des volumes d'annotations utiles pour plusieurs projets actuels (on mentionne 100 M annotations, par exemple, dans [5]), les coûts deviennent absolument rédhibitoires, et forcent chercheurs et organismes subventionnaires à explorer des stratégies plus réalistes.

3 Brève typologie des approches coopératives

Afin de construire les ressources évoquées à la section précédente, on peut recourir à la coopération (*collaborative creation*). Cette coopération peut se faire de façon traditionnelle, par exemple lorsque plusieurs spécialistes collaborent *via* un réseau informatique.

Avec l'avènement d'Internet, des stratégies de collaboration d'une plus large échelle ont vu le jour. Wikipédia en est le meilleur exemple, et fait le pari que l'intelligence collective (*collective intelligence*) d'une foule d'internautes peut rivaliser en qualité avec celle attendue d'experts du domaine. Les études sur Wikipédia semblent valider l'approche [2].

En juin 2006, dans le magazine *Wired* [8], Jeff Howe identifie un nouveau phénomène, le *crowdsourcing* (externalisation ouverte, en français), par lequel des sociétés recourent à un « appel à tous » à de larges groupes de personnes pour effectuer des tâches à une fraction du prix habituellement requis pour ce genre de travail.

On distingue plusieurs formes de *crowdsourcing* [9], [10], qui dépendent avant tout de la motivation de la « foule » qui exécute la tâche externalisée. On les distingue ci-dessous.

3.1 Contributions involontaires

Ces contributions sont faites le plus souvent à l'insu des utilisateurs. Par exemple, dans le cadre du projet reCaptcha [11], les concepteurs de sites web demandent aux internautes de répondre à une question de reconnaissance de caractères (voir Figure 2). L'internaute lit les caractères et les saisit pour prouver qu'il n'est pas un logiciel de pollupostage. Un robot programmé pour faire ceci en serait incapable : la reconnaissance des caractères déformés est encore impossible pour une machine, mais très simple pour un humain.

Ce faisant, l'utilisateur humain contribue à numériser des livres qui sont pour l'heure sous forme d'images. Son effort, si modeste soit-il, permet en effet de reconnaître quelques mots d'un livre dans une archive numérique qui n'attend qu'à être publiée sous forme de texte simple, une fois complétée la reconnaissance de tous ses mots.



Figure 2 – Un exemple de reCaptcha lors de l'inscription sur un site web

3.2 Contributions volontaires bénévoles

Ces contributions par la « foule » des internautes se font spontanément.

Dans Wikipédia, par exemple, les encyclopédistes amateurs versent leurs connaissances et révisent celles des autres. Sur Wikisource^d, des bénévoles numérisent des livres jusque-là sous forme d'image.

Un nouveau type de contribution à des ressources scientifiques s'appelle *sciences citoyennes* (*citizen science*) et permet à tout un chacun de faire une contribution scientifique. À ce chapitre, le site de Zooniverse^e représente ce qu'il y a de mieux en coopération en ligne. Les quelque 850 000 utilisateurs inscrits y sont conviés à contribuer à divers domaines scientifiques, par exemple pour cataloguer et classifier des galaxies, des vocalises de baleines, des cratères lunaires (voir Figure 3), des données climatiques, etc.

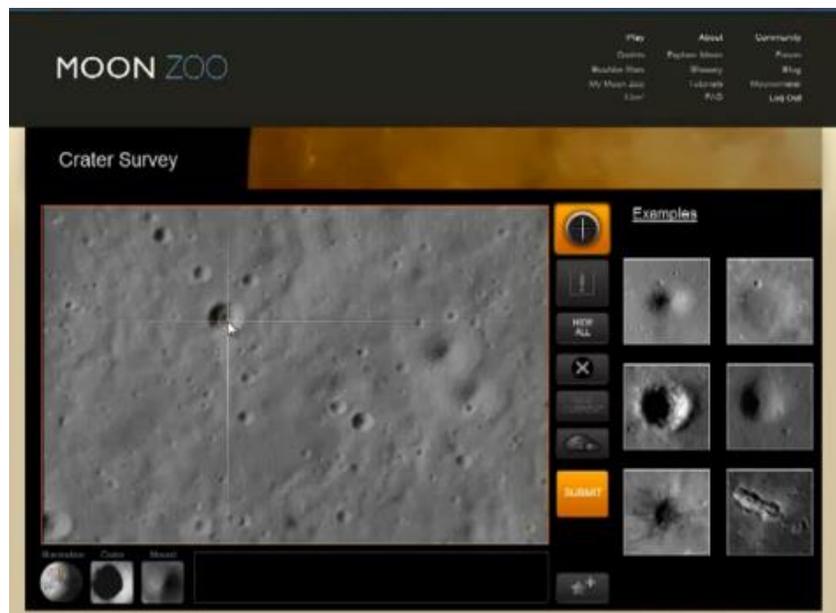


Figure 3 – Interface de classification de cratères lunaires sur Zooniverse.org

3.3 Contributions volontaires rémunérées

La principale plate-forme de *crowdsourcing* en ligne avec rémunération est Amazon Mechanical Turk^f. Elle permet à une organisation de découper une tâche en fragments indépendants et à les soumettre (microtâches) à un très grand nombre de travailleurs en ligne (appelé *turkers*), contre rémunération. Nous donnons plus d'information sur cette approche à la section 8.3.

^d <http://fr.wikisource.org/wiki/Wikisource:Accueil>

^e <https://www.zooniverse.org/>

^f <https://www.mturk.com/mturk/>

3.4 Contributions par le jeu constructif

Le jeu constructif (*game with a purpose* ou *GWAP*) est un jeu (le plus souvent en ligne) qui permet avant tout aux utilisateurs de s'amuser. La seule motivation des usagers est donc le plaisir. Cependant, tout en jouant, les utilisateurs se trouvent à bâtir des ressources d'utilité appréciable, à leur insu ou pas. C'est une façon astucieuse de profiter des centaines de millions d'heures consacrées au jeu vidéo par jour^g.

Le premier jeu constructif d'importance a été le jeu ESP Game, créé par Luis von Ahn [3] en 2003 (voir Figure 4). Deux joueurs y sont invités à tomber d'accord sur un mot qui décrit l'image proposée. Les étiquettes ainsi offertes dans un contexte ludique ont en fait une utilité appréciable : elles servent d'annotations du contenu de l'image (mots-clés), une tâche que les informaticiens savaient difficilement automatiser à l'époque de la conception du jeu^h. En coulisse, les 200 000 joueurs ayant participé ont ainsi permis, depuis la création du jeu la construction d'un corpus de 50 millions de mots-clés anglais pour des centaines de milliers d'images. Google achète une licence du jeu en 2006, et l'utilise dans son Google Image Labeler. Luis von Ahn récidivera notamment avec le jeu Peekaboom, qui attirera 30 000 joueurs, produisant 2 millions d'éléments d'information [13], [14].

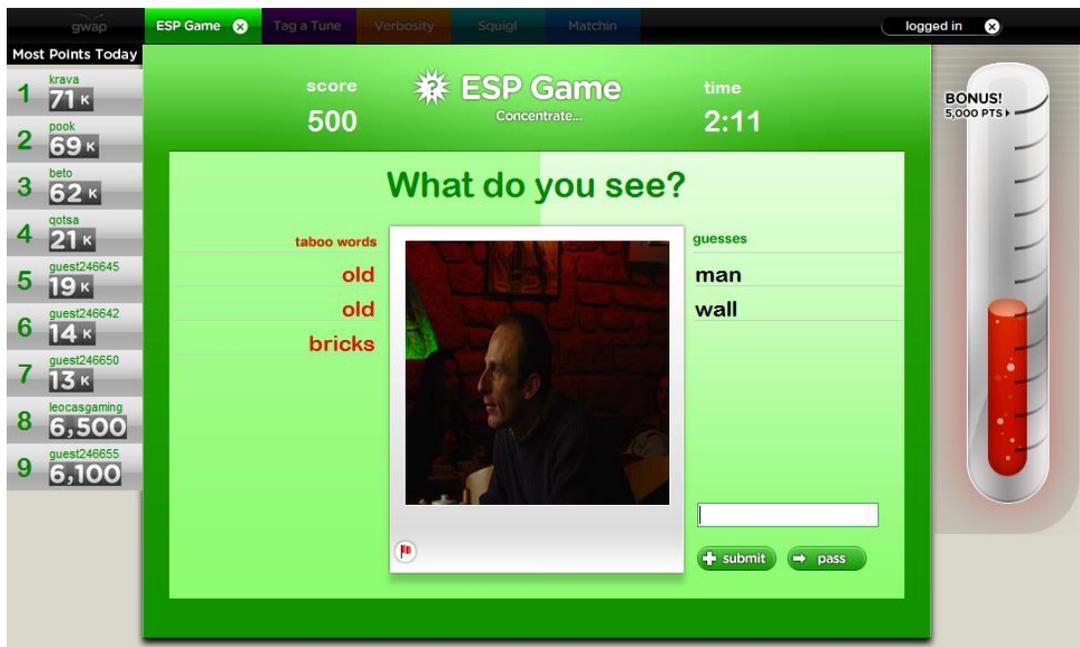


Figure 4 – Jeu ESP Gameⁱ

^g Une étude de la *Entertainment Software Association* rapporte que 200 M heures sont consacrées au jeu vidéo chaque jour aux États-Unis seulement. À l'âge de 21 ans, l'Américain moyen a passé 10 000 heures à jouer à ces jeux, soit l'équivalent de 5 ans à raison de 40 heures par semaine [12].

^h Des progrès considérables ont été faits depuis lors par des méthodes automatiques.

ⁱ <http://www.gwap.com>

4 Quelques distinctions : ludification, jeu sérieux et jeu constructif

Avant de poursuivre avec le jeu constructif, une clarification s'impose sur certains termes utilisés dans le domaine des jeux « non traditionnels ». Ceci nous permettra de rendre plus net l'objectif de ce rapport, et de situer le jeu constructif dans ce paysage.

La « **ludification** » (***gamification***) consiste à emprunter certains des mécanismes du jeu pour les appliquer à des situations non ludiques, afin d'en accroître l'attrait. L'objectif de la ludification n'est donc pas de créer un jeu, mais d'en utiliser certains aspects à d'autres fins. Ces aspects peuvent être des points, des niveaux et des badges, ou de l'argent virtuel de récompense. Par exemple, les utilisateurs du forum de questions-réponses StackOverflow^j se voient attribuer des points et des médailles selon la popularité de leurs questions et réponses sur divers sujets en informatique (voir Figure 5).

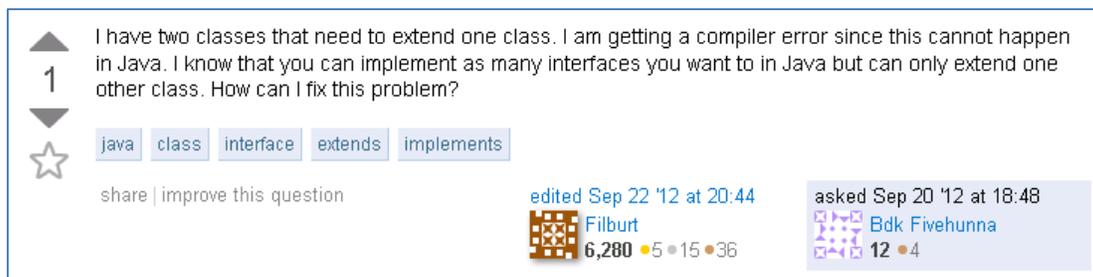


Figure 5 - Exemple de question posée sur www.stackoverflow.com montrant les points et badges des utilisateurs

Le site web The Lottery of Life^k, de l'organisation Save the Children, propose une « roue du destin » où l'on tire au hasard le pays de sa « naissance », afin de sensibiliser l'internaute aux difficultés qu'il aurait rencontrées s'il était né dans ce pays, avec graphiques et sons riches à l'appui (Figure 6). Pourtant, le sujet ne pourrait être plus grave.

Les **jeux sérieux** (***serious games***) appartiennent à un type de logiciel qui combine étroitement un objectif sérieux avec les mécanismes du jeu. Il se doit d'être divertissant, mais permet davantage.

On en compte plusieurs variétés, dont les jeux publicitaires (promouvant une marque, par exemple Chex Quest^l), les jeux d'entraînement (militaire, psychologique, etc.), les jeux favorisant l'exercice physique, la santé mentale, etc.

Le jeu vidéo éducatif est une catégorie importante de jeu sérieux. Il vise à inculquer au joueur des connaissances diverses. On en regroupe diverses variantes, comme le jeu Adibou

^j <http://www.stackoverflow.com>

^k <http://thelotteryoflife.co.uk/>

^l http://en.wikipedia.org/wiki/Chex_Quest

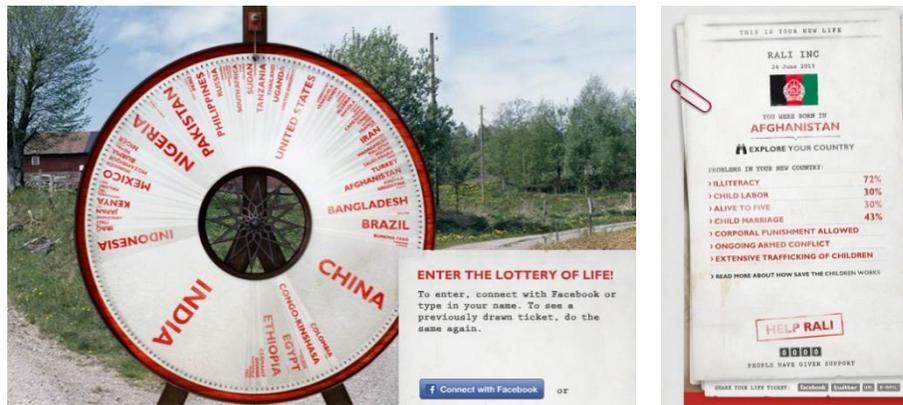


Figure 6 – The Lottery of Life simule la naissance de l'internaute dans un pays du monde choisi au hasard et lui explique les difficultés des enfants dans « son » pays.

pour les enfants de 4 à 8 ans. Les jeux permettant d'apprendre la dactylographie autorisent une combinaison naturelle du jeu vidéo et de l'instruction, comme en témoignent par exemple les jeux Zombie Typocalypse^m ou Word Shootⁿ (Figure 7, page suivante).

Les jeux constructifs sont donc une forme de jeu sérieux, mariant eux aussi la création d'un jeu avec un objectif « sérieux », soit la création de ressources linguistiques.

^m <http://www.wordgames.com/zombie-typocalypse.html>

ⁿ http://cognitivelabs.com/word_shoot.htm



Figure 7 – Dans les jeux Zombie Typocalypse (haut) et Word Shoot (bas), le joueur repousse des hordes d'ennemis en dactylographiant rapidement les étiquettes de ces attaquants afin de les abattre.

5 Deux exemples de jeux constructifs

Cette section présente deux jeux constructifs, un en français, l'autre en anglais, afin d'illustrer notre propos. Nous ferons référence à ces deux logiciels dans le reste de ce rapport, entre autres jeux. L'annexe A propose une liste non exhaustive de jeux constructifs.

5.1 JeuxDeMots

JeuxDeMots^o est un jeu constructif [15] à deux joueurs lancé en 2007 par Mathieu Lafourcade, dont le but est l'acquisition d'un réseau lexico-sémantique. Ce genre de réseau relie les termes d'une langue par des relations sémantiques, telles que synonymie, antonymie, *a* fait partie de *b*, etc. Nous rapportons les résultats de la version française du jeu, qui a été déployé en dix langues, incluant l'anglais, le khmer, l'arabe et le comorien.

Deux joueurs appariés aléatoirement par le jeu voient apparaître le même terme vedette *T* à l'écran et une consigne, qui leur indique de trouver d'autres termes en relation avec le mot *T* affiché (Figure 8). Par exemple, on peut leur demander quel instrument permet l'action affichée, ou des synonymes du mot affiché. La partie dure 1 minute et l'utilisateur ne peut soumettre qu'un nombre limité de réponses, ce qui encourage sa précaution.



Figure 8 – Interface de jeu de JeuxDeMots, pour des idées associées au terme « reportage »

^o <http://www.jeuxdemots.org/>

5.2 Phrase Detectives

Phrase Detectives^p est un jeu constructif créé en 2008 par Jon Chamberlain et ses collaborateurs [19],[5] qui propose à l'utilisateur d'effectuer de la *résolution anaphorique*. En d'autres termes, le joueur se voit proposer un texte dans lequel un passage est souligné, et il doit trouver l'antécédent du passage. Un passage peut être par exemple un pronom, ou une entité.

Ancrée dans la métaphore du détective à l'affût de l'antécédent, l'enquête du joueur prend deux formes. Il doit repérer l'antécédent du référent (jeu « identifiez le coupable » (*name the culprit*), voir Figure 10 à la page suivante) ou valider l'identification faite par un joueur précédent (jeu « conférence de détectives », voir Figure 11). Le jeu étant relativement subtil, les détectives en herbe reçoivent une formation avant de commencer, sous la forme de partie de simulation où un certain score doit être atteint. Les scores sont établis en fonction de la validation avec d'autres joueurs. Les incitatifs sont multiples : scores, palmarès des meilleurs joueurs, petites récompenses en argent.

Une version pour Facebook a été développée en 2011 pour explorer l'intégration du jeu constructif dans les médias sociaux, en particulier les incitatifs sociaux au jeu. La version Facebook ne fonctionne pas au moment de la rédaction de cette étude. Une version pour téléphones intelligents et tablettes est en préparation [5].

Les textes anglais ainsi annotés sont issus de Wikipédia, des livres électroniques libres du Projet Gutenberg^q, et du site textfiles.com. Une version italienne est également disponible.

Comme pour JeuxDeMots, la quantité d'annotations est remarquable. Entre son lancement en 2008 et janvier 2012, 3000 utilisateurs « sérieux » ont abattu 5000 heures de travail, soit 2,5 années-personnes. Le débit moyen d'annotation de l'application est de 450 annotations à l'heure, tous joueurs confondus. Les joueurs ont produit 2,5 M d'annotations (total des annotations « trouvez le coupable » et « conférence de détectives »). Ceci a permis d'annoter 407 documents complets (162 000 mots). Il y a davantage d'annotations que de mots, car plusieurs annotations doivent concorder sur le même mot avant d'être considérées valides, un principe d'agrégation que nous développons en section 7.2.

Les auteurs mettent cet effort en perspective : le corpus Gnome, utilisé jusqu'en 2008 pour étudier la résolution anaphorique, ne contenait que 3000 annotations manuelles. OntoNotes 3.0^r, un vaste projet d'annotation riche mené traditionnellement à grands frais, ne contient « que » 140 000 annotations.

Le but ultime de Phrase Detectives est d'annoter 100 millions de mots, selon ses créateurs [5], ce qui, dans le domaine, est un très grand volume. Les mêmes auteurs estiment qu'une approche traditionnelle d'annotation de la même quantité de texte coûterait plusieurs dizaines de millions de dollars américains.

^p <http://anawiki.essex.ac.uk/phrasedetectives/>

^q <http://www.gutenberg.org/>

^r <http://www ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC2009T24>



USERPROFILE

fbg
36 this week
 2 decisions
 34 agreements
 0 extras

190 this month
190 all time

Level: **Detective**

Your rating: **100%**

CASE OPEN
 47 tasks remaining

2 completed cases

[EDIT PROFILE](#) | [LOGOUT](#)

[Instructions](#)

[FAQ](#)

[Restart](#)

SHARE THIS



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.



Alice in Wonderland (Lewis Carroll)
 Alice turned and came back again.

"Keep your temper," said the Caterpillar.

"Is that all?" said Alice, swallowing down her anger as well as she could.

"No," said the Caterpillar.

It unfolded its arms, took the hookah out of its mouth again, and said, "So you think you're changed, do you?"

"I'm afraid, I am, sir," said Alice. "I can't remember things as I used -- and I do n't keep the same size for ten minutes together!"

"What size do you want to be?" asked the Caterpillar.

"Oh, I'm not particular as to size," Alice hastily replied, "only one does n't like changing so often, you know. I should like to be a little larger, sir, if you would n't mind," said **Alice**. "Three inches is such a wretched height to be."

"It is a very good height indeed!" said the Caterpillar angrily, rearing itself upright as it spoke (it was exactly three inches high).

In a minute or two, the Caterpillar got down off the mushroom and crawled away into the grass, merely remarking, as it went, "One side will make **you** grow taller, and the other side will make you grow shorter."


Not mentioned before

Done


- [▶▶ Skip this one](#)
- [▶▶ Skip - closest phrase can't be selected](#)
- [▶▶ Skip - closest phrase is no longer visible](#)
- [▶▶ Skip - error in the text](#)

SEARCH CLUES

Words like **they, him, her** and **it** are likely to refer to something else in the text. Try to find the closest mention of this 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 previous mention** of the phrase to score maximum agreement points.


Feedback

Figure 10 – Partie de Phrase Detectives, pour l'article Wikipédia *Les Aventures d'Alice au pays des merveilles*, en mode « Identifiez le coupable »

"One side of what? **The other side of what?**" thought Alice to herself.

The phrase in orange **has not** been mentioned before


Disagree

Agree


Figure 11 – Partie de Phrase Detectives, en mode de validation

6 Conception des jeux constructifs

Cette section s'intéresse aux principes de conception des jeux constructifs, c'est-à-dire de leur mécanique ludique et des mécanismes utilisés pour recueillir les données que les utilisateurs génèrent tout en s'amusant.

Les principes de fonctionnement d'un jeu constructif devraient encourager les joueurs à exécuter diligemment les étapes de résolution du problème qui leur est soumis, et, si possible, de fournir une garantie statistique que les ressources ainsi générées sont de qualité. La qualité fait l'objet de la section suivante.

6.1 Trois mécaniques de fonctionnement

Von Ahn [12] distingue trois types de mécanique de jeu constructif. Pour chacune de ces stratégies, des *entrées* (par exemple un mot) sont proposées aux joueurs, et sont l'objet de la construction de la ressource linguistique. Des *sorties* (des annotations) sont produites lors de la partie.

6.1.1 Jeu avec accord sur les sorties (*output-agreement games*)

Deux joueurs sélectionnés aléatoirement se voient présenter la même entrée et les mêmes consignes et sont conviés à produire une sortie. Les participants ne peuvent communiquer et ne peuvent voir les sorties l'un de l'autre. Ils marquent des points lorsqu'ils s'entendent sur les sorties produites. Lorsque l'entente est partielle seulement, un score moindre peut refléter cela.

L'interaction avec un opposant (présumément) humain est un élément ludique. C'est le principe de JeuxDeMots, où les entrées sont le terme vedette et la relation lexicale, et les sorties sont les mots en relation avec le terme vedette.

6.1.2 Jeu avec inversion du problème (*inversion-problem games*)

Le jeu présente une entrée, par exemple un mot, à un joueur A, qui doit le faire deviner à un joueur B, en utilisant différents indices. Les participants, appariés aléatoirement, ne peuvent communiquer autrement que par les indices. Ils gagnent des points lorsque le joueur A fait correctement deviner au joueur B son entrée.

C'est le principe fondateur de jeux télévisés bien connus comme Password, Pyramide, ou Action Réaction. Un exemple de jeu constructif utilisant cette approche est Verbosity^s, créé par Luis von Ahn, illustré à la Figure 12. Le joueur A fait deviner un mot au joueur B, en répondant à des questions posées sur le mot à deviner. Ce faisant, le jeu recueille des axiomes simples de connaissance du monde (*common-sense knowledge*), du type « rhinocéros est un type d'animal » [20], [21]. L'utilité de ce genre de ressource a déjà été

^s <http://www.gwap.com/>



Figure 12 – Partie du jeu constructif Verboosity, tiré de [12]

exposée à la section 2.2. Puisque le nombre d'indices possible est très limité, l'utilisateur qui fait deviner l'autre a tout intérêt à produire des sorties aussi précises que possible.

6.1.3 Jeu avec accord sur les entrées (*input-agreement games*)

Deux joueurs sélectionnés aléatoirement voient s'afficher chacun une entrée. Seul le jeu sait si ces deux entrées sont identiques ou différentes. Les participants doivent produire des sorties qui décrivent leur entrée, et le but de la partie est de deviner s'ils ont des entrées identiques ou différentes. Ici aussi, la mécanique ludique encourage les joueurs à produire des sorties aussi précises que possible. Pour éviter les réponses aléatoires, le système peut encourager les séries de bonnes réponses en offrant des scores proportionnels à la longueur de cette série de succès (et en pénalisant les échecs, par exemple).

Nous ne connaissons pas de jeu constructif pour ressources linguistiques utilisant cette approche, mais Von Ahn cite [12] le jeu TagATune^t, où les compétiteurs devinent s'ils ont en entrée la même chanson, en décrivant le morceau à l'aide de mots. Le jeu collige ainsi des étiquettes décrivant des chansons, fort utiles par la suite pour la recherche de pièces musicales. Cela peut également servir au calcul de la proximité d'un artiste avec un autre, pour des systèmes de recommandation comme Genius[™] sur iTunes[™], d'Apple.

6.1.4 Variantes possibles

Naturellement, il est possible d'envisager la combinaison des approches précédentes, par exemple utiliser un système d'accord sur les entrées pour générer des annotations, validées par la suite dans un jeu fondé sur l'accord sur les sorties [22].

^t <http://www.gwap.com/>

On peut également concevoir des jeux avec plus de deux joueurs, quelle que soit la configuration sélectionnée parmi les trois principales présentées plus tôt. Von Ahn souligne [12] que ceci change profondément la nature du jeu, en ajoutant notamment un élément de compétition là où il n'existait pas, ce qui pourrait déplaire à certains. De plus, cela revient en un certain sens à « gaspiller » des ressources humaines.

6.1.5 Partenaire de jeu artificiel (« bot »)

Les jeux multijoueurs ont un attrait indéniable (section 6.3.3), mais deux ou plusieurs joueurs ne sont pas nécessairement libres en même temps pour une partie donnée, en particulier lorsque le jeu n'est pas encore populaire. Il est donc utile de prévoir un partenaire de jeu artificiel [23]. Ce dernier peut être créé par exemple en enregistrant la partie d'un joueur dans un jeu de type « accord sur les sorties » (section 6.1.1), puis en les proposant ultérieurement à un solitaire. C'est le principe utilisé par JeuxDeMots. Le jeu Categorilla (section 6.3.2), lui, propose des réponses générées automatiquement, sans que les utilisateurs se doutent pour autant qu'ils s'opposent à un robot [23].

6.2 Interaction avec le joueur

Puisque les ressources que l'on souhaite construire sont de nature linguistique, il faut que du texte soit impliqué dans les réponses fournies par le joueur. Tous les jeux que nous avons recensés recourent à un de deux mécanismes possibles : soit l'utilisateur saisit ses réponses librement dans les champs prévus à cet effet, soit il sélectionne dans une liste une réponse appropriée. Certains, comme JeuxDeMots, combinent les deux méthodes.

Lorsque la saisie du texte est libre, il y a un risque évident de fautes de frappe, d'orthographe, etc. Plus pernicieux, des risques de collusion entre joueurs peuvent se manifester s'ils abusent de leur liberté de saisie pour communiquer illicitement. Des mécanismes (voir section 7) doivent être prévus pour mitiger ces risques.

6.3 Incitatifs au jeu constructif

Une chose est sûre à propos du jeu constructif, c'est que les participants doivent y trouver plaisir. Ce qui est moins clair, c'est justement comment concevoir des jeux amusants, propres à attirer le joueur potentiel dans un premier temps, à maintenir son intérêt tout au long des parties, et à le faire revenir régulièrement. Un effort à ce chapitre est particulièrement important, étant donné que les tâches d'annotations de texte ne sont pas reconnues pour être divertissantes pour le commun des mortels.

6.3.1 Quelques théories psychologiques

Ce qui incite une personne à jouer aux jeux vidéo touche à la psychologie, et une revue détaillée de ce sujet ne saurait être le propos de ce travail. On se contentera de citer quelques études faites dans le cadre du jeu constructif, qui éclairent certaines recommandations de conception. L'article de Farias et Brasileiro [24] sur les facteurs d'implication (*engagement factors*) à jouer à des jeux vidéo met en lumière des typologies utiles, issues d'autres études. Le travail de Chamberlain et ses collaborateurs [9] est également riche en enseignements.

Les attraits principaux des jeux peuvent se résumer à trois idées clés, que nous développons dans les trois sous-sections qui suivent :

- **Amour** du jeu, c'est-à-dire le plaisir que le participant y trouve ;
- **Gloire**, c'est-à-dire l'élévation du statut social que le joueur peut tirer de ses conquêtes ludiques ;
- **Argent**, qui correspond aux gains financiers possibles.

Ces trois aspects exercent leur emprise de façon différente sur chaque personne. On verra également que certaines motivations peuvent être contre-productives, dans la mesure où certains compétiteurs en quête de gloire ou d'argent peuvent contribuer de façon malhonnête afin d'accélérer leur progression.

Attrait du jeu selon Koster. Farias et Brasileiro [24] citent une étude de Koster qui voit l'attrait du jeu en termes de 4 catégories : l'esthétisme plaisant, les réactions psychologiques viscérales, l'aspect social du jeu et le plaisir associé à la maîtrise mentale d'un problème renouvelé de façon stimulante.

De fait, la dimension sociale du jeu semble un attrait puissant d'un jeu, à la fois en termes de compétition avec d'autres (palmarès, meilleurs scores, etc.) et de coopération (échanges, coopération). C'est heureux, car les plates-formes sur Internet se prêtent tout naturellement à ce genre d'interactions.

6.3.2 Stratégies incitatives favorisant le plaisir du jeu

Plusieurs études [5], [9], [12], [25] mettent en application diverses théories sur le jeu pour proposer des principes de conception afin de maximiser le plaisir du jeu. Nous les résumons ici, et les enrichissons de nos observations. Toutes ne sont pas applicables à un jeu donné.

Esthétique du jeu. L'attrait graphique est important. Cependant, le jeu constructif s'appuyant sur des éléments linguistiques a forcément recours à des contrôles purement textuels, ce qui ne favorise pas nécessairement des environnements graphiques riches. Des jeux comme la campagne d'évaluation GIVE ont su relever élégamment ce défi [26]. Ainsi, afin d'évaluer des systèmes de génération d'instructions dans un environnement virtuel 3D, les joueurs sont guidés par ces systèmes pour atteindre des objectifs dans les environnements virtuels. L'efficacité des joueurs sert donc de mesure indirecte de la qualité du système de génération de texte. L'interface est illustrée à la Figure 13.

L'interface du jeu Wordrobe^u est un autre exemple d'interface graphique très léchée sur Internet, et un modèle dans le genre, illustrée à la Figure 14.

Score et progression. Un score est à la fois naturel et indispensable pour que l'utilisateur puisse apprécier son progrès dans le jeu, et, le cas échéant, pour se comparer aux autres. Le calcul du score va du très simple (un point par contribution), jusqu'au plus complexe, faisant intervenir la qualité perçue de l'annotation, par exemple en mesurant l'accord avec un autre annotateur, comme dans JeuxDeMots.

^u <http://wordrobe.housing.rug.nl/>



Figure 13 – Interface de jeu GIVE

Réponse en un temps limité. Ceci introduit un élément de défi, et, lorsque la difficulté est bien calibrée, renforce le sentiment de maîtrise mentale d'un problème intéressant. Cela maximise aussi le débit d'annotation (*throughput*) et la spontanéité des réponses, ce qui peut être désirable. La technique est utilisée par la plupart des jeux constructifs (JeuxDeMots et Verbosity). Parmi les exceptions notables, Phrase Detectives présente des parties sans chronomètre, après que les testeurs se sont plaints de se sentir « sous pression » [9].

Concentration. Les jeux constructifs devraient proposer une interface et un problème favorisant la concentration, comme les mots croisés. Ceci favorise une immersion dans le jeu qui amplifie le plaisir d'y participer et d'en franchir les épreuves.

Feedback. Liée à la question du score, les participants devraient idéalement avoir une rétroaction positive lorsque leur contribution s'avère utile, abondante, ou persistante. Le

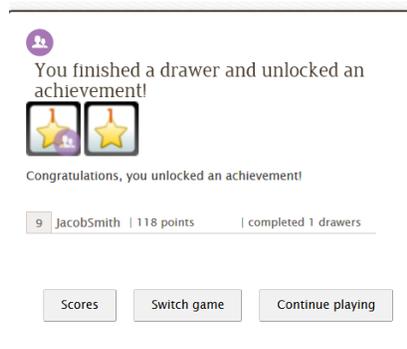


Figure 14 – Exemple de rétroaction positive pour le jeu Wordrobe

score est un moyen évident d'atteindre cela, mais on peut y coupler des récompenses, badges, ou autres attestations de l'effort investi. La collection de jeux constructifs Wordrobe est un exemple très bien fait à cet égard (Figure 14), permettant d'accumuler des badges (*achievements*) de plusieurs types [27]. Il faut cependant noter que, dans plusieurs architectures, le système *ne sait pas quelle est la bonne réponse* (c'est justement ce qu'il cherche à connaître), ce qui complexifie la rétroaction.

Niveaux de difficulté. Les niveaux de difficulté permettent à l'utilisateur de mesurer sa progression dans le jeu, renforcent le sentiment de contrôle sur le jeu et définissent des objectifs clairs au jeu. Si le jeu se prête à moduler sa difficulté d'annotation, ces niveaux permettent également de garder le niveau de difficulté en phase avec les capacités grandissantes du joueur, pour maintenir son intérêt. À la fin d'une partie, par exemple, l'interface montre au joueur où il est en est dans sa progression, et le nombre de points nécessaires pour passer au niveau supérieur. Si un jeu doit appairer aléatoirement deux joueurs, il pourrait le faire soit en forçant un niveau de difficulté similaire, soit avec un petit écart, pour que le joueur plus faible soit encouragé à s'améliorer.

Dans le contexte de la cueillette de ressources linguistiques, cela pourrait être utile pour mesurer la fiabilité des contributions d'un participant donné.

Dans Phrase Detectives, fidèle à la métaphore du travail de détective, les joueurs passent du stade de « recrue » à celui de « détective », pour souligner leur effort (Figure 15).

Les concepteurs du jeu Sentiment Quiz, qui s'intéressent à la polarité d'une phrase (sentiment positif ou négatif), considèrent même la création de « tournois de champions » pour ceux qui auraient atteint un niveau donné [22].

Éléments aléatoires. Les jeux constructifs doivent nécessairement inclure des éléments choisis aléatoirement, ne serait-ce que pour proposer une unité à annoter choisie au hasard. On l'a vu plus haut, l'appariement de deux joueurs doit également se faire au hasard, pour éviter la collusion dans les jeux où l'accord entre les deux participants détermine le score.

Imitation de jeux existants. Vickrey et ses collaborateurs [23] proposent bien justement que si les gens aiment un jeu de société dans la « vraie vie », ils pourraient aussi aimer prendre part à des jeux constructifs qui leur ressemblent. Ces chercheurs se sont inspirés de Scattergories™ pour créer Categorilla, un jeu construisant un réseau lexical (comme le



Figure 15 – Progression du niveau « recrue » à « détective » dans Phrase Detectives

réseau bâti par JeuxDeMots).

Scattergories™ et Categorilla demandent aux utilisateurs de trouver des mots ou des expressions qui conviennent à certaines catégories, comme « des choses qui volent » ou « des types de poissons ». De plus, les réponses doivent commencer par une lettre donnée, choisie au hasard. Les deux jeux sont illustrés à la Figure 16. Les auteurs récidivent avec leur logiciel Free Association, inspiré de Taboo™, où les participants ajoutent des mots en relation avec un autre, en évitant une liste fixe de mots « tabous ».

6.3.3 Stratégies incitatives liées aux aspects sociaux du jeu

Compétition. La compétition entre joueurs en ligne permet un certain défi attrayant supplémentaire, permettant de mesurer ses forces contre celles d'un opposant, souvent choisi au hasard. Notons cependant que, pour diverses raisons, ce ne sont pas tous qui ont le goût de la compétition, même ludique. Ces derniers s'amuseront davantage en solo.

Coopération. La collaboration permet à deux participants ou plus de s'entraider en vue d'atteindre les objectifs fixés par le jeu. À notre connaissance, aucun jeu constructif ne met en œuvre ce principe, cependant. Il y a, à notre avis, un potentiel inexploré ici.

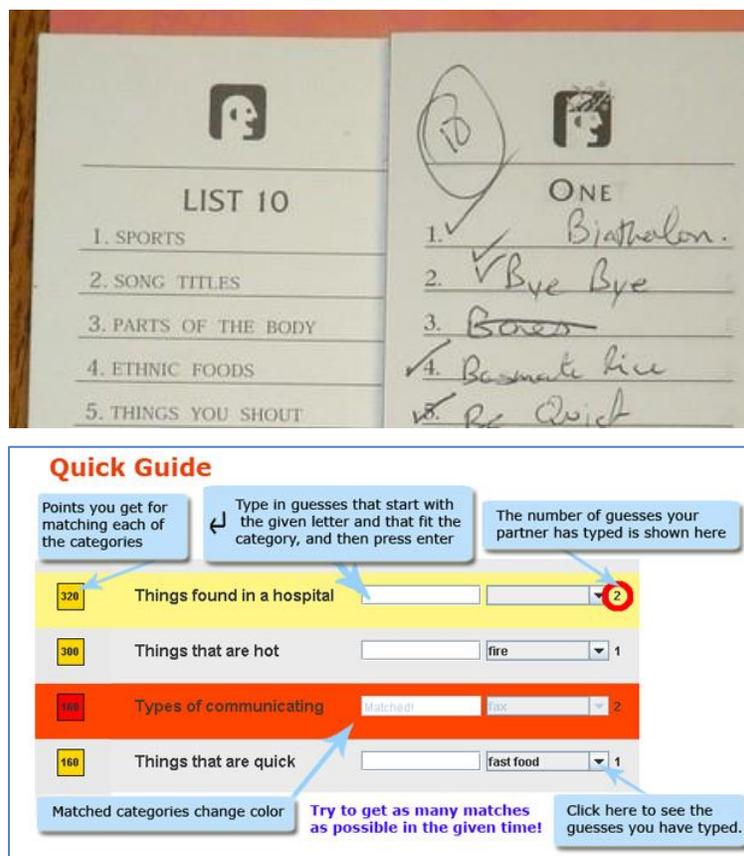


Figure 16 - Partie du jeu de société Scattergories™ pour la lettre « B » (haut) et interface de tutoriel de Categorilla (bas)

Palmarès des meilleurs scores. La plupart des jeux proposent un tableau d'honneur indiquant les meilleurs scores des internautes y ayant participé. Ceci recoupe l'élément incitatif « gloire » mentionné en section, soit le plaisir d'afficher ses succès dans son réseau social. Un élément d'historique est aussi souvent présent, permettant de consulter les meilleurs scores sur une période bien déterminée.

Les médias sociaux se prêtent bien à cet étalage, et le jeu constructif en profitera donc pleinement. La plate-forme de jeu Wordrobe est un modèle du genre (Figure 17), permettant de filtrer les résultats selon chacun des jeux constructifs qu'elle propose, et de consulter le profil Facebook des meilleurs.

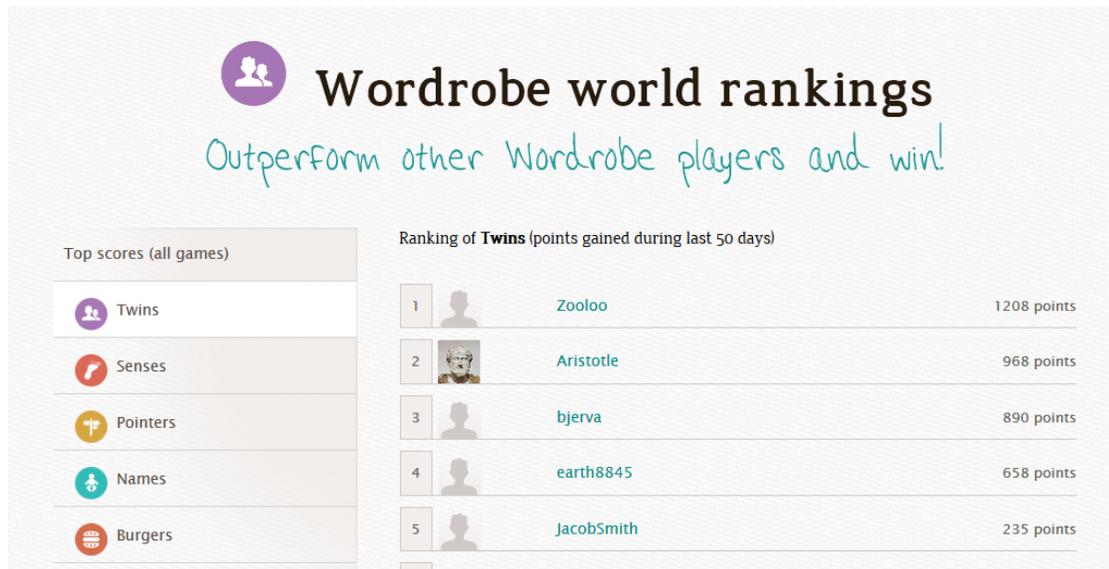


Figure 17 – Palmarès des meilleurs joueurs dans Wordrobe

Altruisme. Une stratégie liée à l'aspect social du jeu est de profiter du sens altruiste du joueur intéressé à contribuer à sa façon aux poursuites scientifiques, ce que l'on a appelé les sciences citoyennes à la section 3.2. De fait, on peut lire dans un blogue portant notamment sur le jeu Verbosity que, même si ce dernier est « de l'entrée de données glorifiée », il est addictif en plus d'être « incroyablement utile pour rendre les moteurs de recherche plus efficaces »^v. L'étude de Farias [24] recommande fortement cette façon de faire, car leur étude sur le profil des participants à leur jeu constructif montre que plus les objectifs scientifiques sont clairs, plus les participants comptent de « non-joueurs » intéressés simplement à épauler les chercheurs.

6.3.4 Incitatifs financiers : un pari risqué ?

Ce travail s'intéresse notamment au jeu constructif auquel participent des volontaires non rémunérés. Cependant, il est naturel d'envisager l'octroi d'une somme d'argent relativement modeste (par rapport aux autres coûts d'implémentation) pour attirer davantage de

^v Notre traduction d'un extrait de <http://www.emmaroseclarke.com/1/archives/01-2009/1.html>

compétiteurs. Ces sommes étant faibles, elles ont davantage une valeur symbolique, et peuvent servir autant à attirer de nouveaux adeptes qu'à encourager les contributions des convertis de longue date.

C'est une stratégie qu'ont notamment employée les concepteurs de Phrase Detectives et Wordrobe. Les concepteurs de la collection de jeux Wordrobe ont récompensé les meilleurs joueurs, à raison de 15 £^w, 20 £ ou 25 £ en fonction du jeu où ils s'illustrent, selon une entrée de leur page Facebook^x. Pendant quelques mois, les concepteurs ont également offert des bons d'échanges Amazon, pour une valeur totale de 150 £ (240 CAD). Les effets de ces incitatifs n'ont pas été publiés.

L'équipe de Phrase Detectives a récompensé les meilleurs joueurs, à raison de 5–10 £ par jour, 10–15 £ par semaine et 30–75 £ par mois. De plus, des joueurs tirés au hasard se voyaient également remettre un prix, créant un effet de loterie [5]. Les concepteurs ont constaté une augmentation du nombre d'annotations produites, attribuable à l'augmentation du volume d'annotations, et non pas à un plus grand nombre de participants actifs [9].

Cependant, les incitatifs financiers peuvent s'avérer contre-productifs, s'ils induisent certains joueurs sans scrupules à produire un grand volume d'annotations de piètre qualité. C'est un problème déjà observé par certains utilisateurs de Mechanical Turk [22]. Par exemple, une étude en question-réponse [28] a montré que l'augmentation du montant de la récompense accélère le processus d'annotation au détriment de la qualité. Ces abus font l'objet de la section 7.4.

6.4 Jeu constructif et médias sociaux

Nous l'avons évoqué plus haut, les médias sociaux se prêtent bien à la mise en œuvre de jeux constructifs. Facebook est la plate-forme la plus populaire à cet égard. Cette dernière présente plusieurs qualités désirables. Le jeu constructif Sentiment Quiz [22] résume à lui seul presque tous les avantages de la populaire plate-forme. Le jeu consiste à deviner si un extrait de texte est positif ou négatif, et a servi à étudier la perception de l'information diffusée dans les médias (traditionnels et autres) lors des élections présidentielles américaines de 2008 (voir Figure 18).

Compétition amicale. Les utilisateurs peuvent entrer en compétition entre eux, et peuvent se comparer les uns aux autres. Dans Sentiment Quiz, ces derniers se « voient » sur un axe graphique horizontal, et un palmarès des champions attise la rivalité ludique entre abonnés Facebook (Figure 18).

Collaboration. Facebook pourrait se prêter à des efforts collaboratifs, mais aucun jeu constructif recensé ne profite de cette (complexe) possibilité.

Récompenses. Des récompenses et badges peuvent être affichés sur le fil (*timeline*) de l'abonné Facebook, au vu et au su de ses contacts.

^w 1 £ = 1,58 CAD en date de la rédaction de ce rapport.

^x <https://www.facebook.com/Wordrobe>

facebook Friends Applications Inbox Home Search

US08 Sentiment Quiz
Game About Awards Feedback

ECOresearch.net

Sentiment Quiz Awards Hall of Fame

Does this sentence use positive or negative language?

“ We are headed down a path that is certain to end in the destruction of our experiment in democracy. ”

Hover over the buttons to see what they stand for.

Status

Level 5

Your current score is 127 points.
Inviting 1 friend has earned you 21 points.

Spread the Word

Tell your Friends! You will earn 10% of your friends' points after they accept your invitation! The calculation is recursive, so if they invite others you will even get more bonus points.

Rank	Name	Score
1.	Walter	535
2.	Julia	361
3.	Hermann	211
4.	Reinhard	205
5.	Norberto	204
...		
8.	Fábio	154
9.	Isabella	130
10.	You	127
11.	Lord	121
12.	Sandra	102

Tips & Tricks

Use the Report Button if you encounter quotes that contain strange characters or make absolutely no sense. You can get points for that, too!

The Neutral Button is meant for quotes which appear to have no sentiment at all, for example **Hillary is on her way to New York.**

New Media MBA
www.modul.ac.at/nmt/mba

EDITED BOOK
The Geospatial Web
Geobrowsers, Social Software & the Web 2.0

Page built by Sentiment Quiz (report) About Find Friends Advertising Developers Terms Privacy Help

Figure 18 – Interface du jeu Sentiment Quiz, tirée de [22]

Publicité. Le marketing viral est une des grandes forces de Facebook. Un jeu peut donc profiter d'une visibilité très accrue si le joueur « aime » le jeu, ou si l'application contribue à peupler le fil (*timeline*) du joueur.

Sentiment Quiz ajoute à cela une fonctionnalité bien connue en marketing, soit la possibilité d' enrôler un ou plusieurs amis afin de profiter d'une récompense dans le jeu. Le participant rusé peut ainsi profiter de 10 % des points que ses amis recueillent (qui peuvent à leur tour inclure des points gagnés par invitation).

Plate-forme de développement. Détail technique non négligeable, Facebook a une plate-forme de développement prête à l'emploi, à même de simplifier le développement, le déploiement et le soutien technique pour les informaticiens impliqués. Le réseau social permet également une certaine stabilité, puisque les concepteurs n'ont pas à maintenir un serveur fonctionnel par eux-mêmes ; l'hébergement est assuré par Facebook.

7 Qualité des annotations

Le jeu constructif a comme raison d'être la création d'annotations de qualité. Cet objectif n'est pas évident, considérant que :

- les annotateurs sont par définition des non-experts du domaine ;
- les annotateurs commettent des erreurs d'inattention ;
- les joueurs sont parfois davantage motivés par le gain de points et de récompenses que par la qualité du travail fourni, présumément plus s'ils ne sont pas conscients de l'aspect scientifique de leur activité ;
- le processus d'annotation lui-même peut s'avérer très complexe, soit parce que la tâche est difficile par elle-même (par exemple, la résolution anaphorique), soit parce qu'une partie donnée est plus difficile.

Il est bon d'insister sur le dernier point : plusieurs tâches en TALN ont une part d'ambiguïté qui peut défier même les experts humains. Par exemple, la phrase « Alice voit Bob avec un télescope. » admet deux interprétations.

Nous proposons dans cette section de considérer les métriques permettant d'évaluer la qualité des annotations, ainsi que plusieurs recommandations permettant de l'encourager.

7.1 Métriques de qualité

La qualité des annotations est habituellement évaluée en les comparant à des annotations de référence préparées par des experts. Les ambiguïtés de certaines tâches bornent supérieurement la performance de ceux-ci. Un score de 0 % indique un accord nul entre une référence et les annotations produites, un score de 100 % indique un accord parfait.

Phrase Detectives. Les accords entre deux annotateurs humains et les sorties du jeu sont présentés au Tableau 2, pour un échantillon de 5 documents annotés. Les experts humains s'entendent bien entre eux (ce qui indique une méthodologie saine), et fixent une borne supérieure de 94,1 % d'accord pour le jeu.

L'accord entre les experts et le jeu est de l'ordre de 84 %, ce qui est comparable à la qualité produite traditionnellement par des annotateurs entraînés à la tâche par des experts (voir section 2.3).

Tableau 2 - Accord sur les annotations de Phrase Detectives, par type, tiré de [5]

Type d'annotation	Expert 1 & Expert 2	Expert 1 & jeu	Expert 2 & jeu
Général	94,1 %	84,5 %	83,9 %
Apparition d'une nouvelle entité	93,9 %	96,0 %	93,1 %
Référence à un antécédent	93,3 %	72,7 %	70,0 %
Impersonnel	100,0 %	100,0 %	100,0 %
Propriété d'une entité déjà mentionnée	100,0 %	0,0 %	0,0 %



Figure 19 – Jeu AKI (<http://www.jeuxdemots.org/AKI.php>)

La difficulté de la tâche a un effet sensible sur la qualité des annotations. Déterminer qu'une entité est impersonnelle (p.ex. le mot « il » dans « il pleut ») semble donner de bons résultats. Par contre, déterminer qu'un extrait de texte fait référence à une propriété d'une entité déjà mentionnée dans le texte semble très difficile par les participants, et se traduit par un accord nul (dernière ligne du Tableau 2).

JeuxDeMots. L'évaluation de JeuxDeMots est plus complexe, car il n'y a pas de ressource de référence pour évaluer le genre de réseau lexico-sémantique construit librement par les joueurs. Vu la taille de la ressource, une évaluation par un expert est prohibitive. Les concepteurs de JeuxDeMots ont donc (comme il se doit...) conçu un jeu constructif appelé AKI permettant l'évaluation du réseau [29]. Le jeu, illustré à la Figure 19, demande à l'utilisateur de penser à un mot et de fournir des indices pour que le jeu le devine. AKI se sert alors du réseau lexical pour converger vers le mot à deviner.

AKI devine justement dans environ 80 % des cas, ce qui fait conclure aux auteurs que 75 % des termes du réseau lexical sont ainsi correctement indexés.

7.2 Agrégation des annotations et validation

La plupart des auteurs n'utilisent les annotations produites par le jeu constructif que lorsqu'elles font l'objet d'un accord entre une ou plusieurs personnes. En agrégeant ainsi ces données, on en augmente la fiabilité. Ceci peut se faire de diverses façons, en tenant compte

de la nature des ressources colligées. Par exemple, on peut recourir aux méthodes suivantes pour valider une annotation donnée :

- unanimité entre plusieurs joueurs ;
- vote de majorité absolue entre tous les joueurs s'étant prononcés sur l'unité à annoter ;
- vote de majorité relative entre tous les joueurs s'étant prononcés sur l'unité à annoter ;
- test du χ^2 où l'hypothèse nulle est une distribution des réponses qui est identique à une distribution uniforme [27];
- vérification par un expert (plus rapide que la *production* de l'annotation par un expert) ou vérification par plusieurs niveaux d'experts ;
- moyennes.

Notons cependant que même une annotation faite une seule fois peut fort bien être valable. À l'inverse, une annotation où il y a unanimité peut être erronée.

Une étude [30] propose une formule pour calculer le nombre de fois qu'un même extrait doit être annoté par un joueur pour obtenir un consensus sur l'annotation. Ce nombre de répétitions est $(\text{nombre d'annotations pour un consensus} - 1) \times (\text{nombre de réponses possibles}) + 1$. Le nombre d'annotations pour un consensus est un nombre arbitraire décidé par les concepteurs de la tâche. Une étude [9] suggère une valeur entre 5 et 10. Ces stratégies avec agrégation réduisent (parfois significativement) le nombre d'annotations valables. Pour le jeu Categorilla [23], deux tiers des annotations sont supprimés.

Une autre approche est le concept de joueurs de validation, à qui l'on présente une annotation et qui doivent se prononcer sur sa validité. C'est la stratégie adoptée par Phrase Detectives, lorsque le participant est en mode « validation » (Figure 11). Elle a elle aussi ses faiblesses : si une annotation ne fait pas l'unanimité, qui a raison ? De plus, cette mésentente peut être attribuable à un passage particulièrement ambigu, où des annotations contradictoires en surface sont en fait toutes légitimes. Il peut y avoir un bogue dans la préparation des données, également, se traduisant par une partie impossible à gagner. Il n'y a pas de panacée, ici.

7.3 Principes de conception pour des annotations de qualité

Interface graphique. Comme tout logiciel, une interface utilisateur bien conçue favorisera un usage responsable du jeu constructif. Le problème est d'autant plus délicat que les jeux permettent rarement de revenir en arrière lorsqu'une erreur d'inattention est commise : cela rendrait le défi trivial et inintéressant.

En général, l'interface doit être simple et claire, construite avec le public cible en tête, et doit être d'un apprentissage intuitif [5].

Conception de la tâche. La tâche d'annotation doit être conçue de façon à encourager les participations responsables. La méthode la plus simple est de fournir des choix de réponse lorsque c'est possible. Lorsque la nature des annotations ne s'y prête pas, il faut être créatif.



Figure 20 – Jeu constructif 1001 Paraphrases pour la reformulation de « this can help you »

Les concepteurs du jeu 1001 Paraphrases, qui collige des reformulations d’extraits de phrases, se sont intéressés à la conception d’un jeu constructif encourageant la rigueur [25]. Ainsi, leur interface de jeu (Figure 20), plutôt que de demander aux participants de saisir librement des reformulations, les invite à compléter des réponses partielles.

L’idée est applicable à une question ayant plusieurs bonnes réponses. Pour chaque question, l’approche nécessite au minimum une bonne réponse « initiale », recueillie par exemple dans un autre jeu, ou par un processus assurant une excellente qualité de la réponse. Une fois générée, cette réponse initiale est proposée aux joueurs après censure partielle, afin d’aiguiller leurs contributions. Les auteurs n’élaborent pas sur le risque de biais que ce genre de réponse-amorce introduit dans les réponses subséquentes.

Possibilité d’indices. Quand la tâche d’annotation s’y prête, des indices peuvent être fournis à l’utilisateur. Pour 1001 Paraphrases mentionné au paragraphe précédent, un bouton « indice » (Figure 20) révèle une partie de la réponse, mais le joueur verra son score amputé. L’implémentation de ce genre de stratégie reste limitée, à notre avis. Le recours à ces indices sert également, par effet de bord, à mesurer la difficulté d’une partie du jeu et à profiler un utilisateur qui abuserait des indices (voir « profilage » ci-dessous).

Possibilité de passer une partie. Plutôt que de forcer un joueur à se prononcer sur une partie qui le confond, plusieurs jeux permettent de « passer une partie » (voir le bouton « Give up » à la Figure 20, par exemple). Cela évite à l’utilisateur de saisir n’importe quoi pour aller de l’avant.

Certains jeux, comme JeuxDeMots et PhraseDetectives, vont plus loin et permettent à l’utilisateur de signaler qu’il n’y a pas de bonne réponse possible, ou qu’un problème dans le

texte empêche toute réponse. Par exemple, dans JeuxDeMots, l'instruction indique « Si vous ne savez pas répondre, il faut passer la partie. Si vous pensez qu'il n'y a pas de réponse possible, vous pouvez mettre *** ». Si le joueur passe ou ne saisit rien, JeuxDeMots propose parfois de signaler au jeu qu'un mot est mal orthographié. Le bas de la Figure 10 montre les options similaires offertes par Phrase Detectives.

Formation de l'utilisateur. L'excellente étude de Chamberlain [9] met en garde contre l'erreur de faire tester le jeu constructif par des collègues et amis : ce n'est pas le public cible, qui pourra être confondu, lui, par la difficulté de la tâche. Il faut au contraire s'assurer que le jeu est bien expliqué, et qu'un tutoriel soit disponible, le cas échéant. Ce dernier doit être adapté au niveau de compréhension du participant moyen.

Phrase Detectives est un bon exemple, et fait commencer l'utilisateur en mode « recrue », pour ensuite le promouvoir si ses réponses correspondent à celles du solutionnaire connu par le jeu. Certains jeux classiques très populaires, comme le viral Candy Crush Saga, ponctuent le jeu de remarques et instructions utiles à l'utilisateur pour lui faire comprendre la mécanique ludique. La Figure 16 (bas) montre l'image d'aide pour le jeu Categorilla. Il faut cependant garder à l'esprit que peu d'utilisateurs lisent ces tutoriels et manuels d'instructions [25].

Réponses taboues. Afin d'encourager la créativité dans les réponses des participants, il est parfois utile de prévoir une liste de réponses taboues, que le joueur ne peut donner, ou pour lesquelles il ne reçoit pas de points. Ces réponses peuvent provenir de multiples parties précédentes (et sont donc ainsi validées) ou d'un prétraitement automatique des données. Dans ce dernier cas, les réponses déjà évidentes pour un programme informatique sont générées, et il est contre-productif pour l'annotateur-joueur de les proposer de nouveau. Ces réponses taboues ajoutent également un élément de défi. On peut ainsi moduler la difficulté d'un jeu, d'un niveau du jeu ou d'une partie donnée en allongeant ou rétrécissant la liste de ces réponses interdites.

Le jeu Free Association (page 21) repose entièrement sur le principe de réponses taboues.

Profilage des utilisateurs. Il est possible de construire un « profil » de joueur permettant de quantifier la qualité de sa contribution. Par exemple, en parsemant les parties normales de parties « truquées » dont le système connaît la réponse, on peut mesurer le sérieux et la compétence d'un joueur [22], [31]. Ses contributions peuvent alors être écartées, ou leur importance pondérée. Nous l'avons dit plus haut, on peut également mesurer le recours systématique à des indices par certains, ou encore utiliser simplement leur niveau de jeu pour les profiler.

Vérification orthographique. Lorsque les entrées sont libres, il y a un risque évident de fautes d'orthographe. Les jeux JeuxDeMots et Verbosity signalent le problème au participant, et JeuxDeMots va jusqu'à proposer des corrections au mot saisi. Pour le jeu 1001 Paraphrases, une analyse manuelle des reformulations de phrases saisies indique que 5 % d'entre elles contiennent au moins une faute. Un échantillon de réponses du jeu Verbosity indique 29 % de réponses agrammaticales [21].

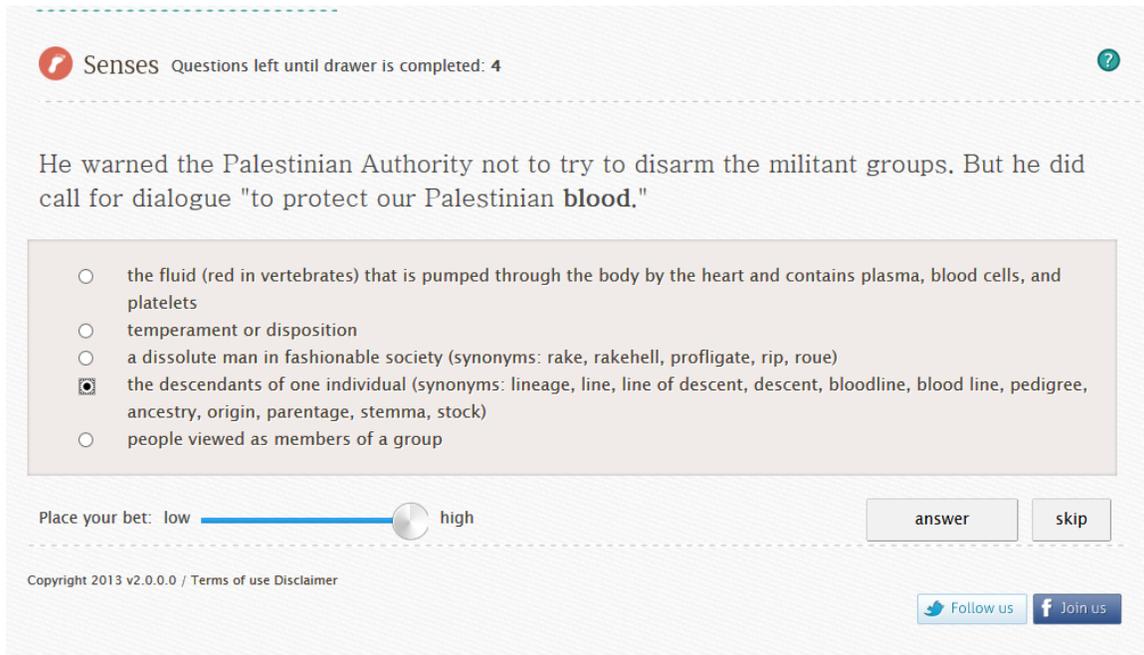


Figure 21 – Interface du jeu *Senses* de la famille de jeux constructifs *Wordrobe*

Niveau de confiance. Il est possible de demander au joueur quelle confiance il place en sa réponse, au moment de jouer. La famille de jeux *Wordrobe* permet ainsi au joueur de « parier » le nombre de points qu’il pourra gagner en répondant correctement, sur une échelle de 0 à 100 (Figure 21). Les chercheurs indiquent que la méthode semble modestement porter fruit [27].

7.4 Principes de conception contre la collusion et les abus

Absence de patron exploitable dans les réponses. Il faut évidemment éviter que le jeu présente une régularité dans les réponses qui puisse permettre de les prévoir sans avoir à résoudre le défi présenté.

Limites sur la communication entre joueurs. Pour empêcher la collusion, il est utile que les jeux multijoueurs mettent en compétition des partenaires qui ne se connaissent pas, en les choisissant aléatoirement, par exemple, et en masquant leur identité.

Ceci n’est cependant pas toujours suffisant. Pour le jeu *Verbosity*, par exemple, des joueurs mal intentionnés détournent les champs prévus pour la description du terme à deviner pour décrire la phonétique ou la morphologie du terme. Dans le panneau de droite illustré à la Figure 22, par exemple, le mot à deviner « headquarters » est décrit par les indices « head quarters » et « joignez ces mots », au lieu de s’en reporter aux indices de sens demandés par les règles du jeu. D’autres indices illicites incluent « rime avec le mot ... » ou encore « commence avec la lettre ... ».



Figure 22 – Tricherie dans le jeu Verbose commise par le partenaire (panneau de droite)

Filtrage des données recueillies. Il est possible de recourir à des filtres en aval de l'étape de collection des données, afin de les assainir avant utilisation. Des heuristiques de tout acabit existent. Conscients du problème illustré à la Figure 22, des chercheurs se sont penchés sur le nettoyage des données du jeu Verbose [21]. Ils ont utilisé comme filtres :

- des mots dans une liste noire pour exclure des réponses (« rime », « commence », « contraire », etc.) ;
- des critères de similarité morphologique entre un indice et la réponse attendue (par exemple le mot « wander » est utilisé illicitement pour faire deviner le mot similaire « wonder ») ;
- des critères de similarité phonétique entre un indice et la réponse attendue (par exemple entre « feat » et « meat ») ;
- diverses heuristiques.

Ces filtres ont éliminé 50 % des contributions des joueurs, soit 250 k sur les 500 k contributions « brutes ».

Analyse de la distribution des réponses. Une simple analyse statistique peut également permettre de détecter les contributions frauduleuses ou négligentes. Les auteurs de Phrase Detectives, par exemple, chronomètrent la vitesse d'annotation et mesurent la distribution des annotations possibles dans les sorties d'un joueur. Ils ont ainsi pu prendre sur le fait un utilisateur qui prenait 4 s pour annoter (la moyenne est de 7 s) et qui répondait la même chose dans 99 % des cas. Le fâcheux a fini par avouer l'utilisation d'un logiciel de complétion automatique de formulaire [9]. Son châtime n'a pas été précisé.

8 Approches coopératives non ludiques

À la section 3, on a distingué plusieurs types d'approches coopératives en construction de ressources linguistiques. Les approches ludiques ont fait l'objet des sections qui précèdent, et nous abordons ici les stratégies de construction coopérative non ludiques, qu'elles soient bénévoles ou rémunérées. Une partie de ce qui a été dit sur les jeux constructifs s'applique également à ces approches non ludiques, et nous inviterons alors le lecteur à se référer aux sections pertinentes.

Rappelons enfin que cette étude s'intéresse à la création de *ressources linguistiques*, décrites à la section 2.

8.1 Motivation des utilisateurs

Comme expliqué aux sections 3.2 et 3.3, les motivations des utilisateurs sont ici l'altruisme (sciences citoyennes) ou une rémunération (plate-forme de microtâches comme Amazon Mechanical Turk). Ceci contraste avec l'aspect de détente et de plaisir des jeux constructifs.

Les contributeurs ayant des motivations différentes, le travail de promotion est différent : au lieu de publiciser la valeur de divertissement auprès des participants potentiels aux jeux constructifs, la clientèle cible est constituée soit de bénévoles intéressés par le sujet et la cause qu'ils appuient, soit par l'appât du gain.

8.2 Exemples d'approches coopératives non ludiques bénévoles

Wiktionnaire^y. Le wiktionnaire est un projet coopératif basé sur une interface de type wiki consacrée à la création d'un dictionnaire libre et multilingue. Ses 2,3 millions d'articles décrivent les mots de plus de 3000 langues, et contiennent définitions, prononciations, étymologies, exemples, antonymes, etc. Son utilité n'est plus à démontrer en traitement automatique des langues (TALN), avec plus de 150 études scientifiques l'ayant employé d'une façon ou d'une autre, selon les archives du domaine^z.

Projet Papillon. Le projet Papillon [7] a pour but la création d'une grande base de données lexicales dans 8 langues (Figure 26), *via* une interface coopérative de type wiki sur Internet. Lancé en 2000 par Emmanuel Planas, François Brown de Colstoun et Mutsuko Tomokiyo, le projet profite d'un partenariat avec le National Institute of Informatics à Tokyo.

Son mécanisme est original. Le projet démarre par une « amorce », première base lexicale contenant de nombreuses entrées, mais peu renseignées. Les contributeurs, spécialistes ou pas, peuvent alors modifier les données existantes. Ces données sont ensuite soumises à l'accord des utilisateurs, spécialistes ou pas. Enfin, l'intégration fait appel à des utilisateurs de confiance pour véritablement verser les contributions (validées ou non) à la base de données. Cette stratégie permet un contrôle serré de la qualité des contributions, mais pose

^y <https://fr.wiktionary.org/>

^z <http://www.aclweb.org/anthology-new/>

FeV			
papillon /PAPIYON-/			
[n.m.]			
(insecte)	butterfly	con bướm	(nghĩa bóng) người nhẹ dạ; người thích phù hoa (thể dục, thể thao)
kiểu bơi bướm	rama-rama	Puteri : I've changed as you wanted, but it's the opposite	
(autocollant, feuillet)	sticker	người chọc; dao chọc (tiết lợn...), gai, ngành	người dán, nhãn
cổ sấn cón dính, nhãn dính có hình	(thông tục) người kiên trì; người bám dai	pelekat	
(contravention)	nhảy (cần quần áo), bướm đêm, sâu bướm	surat denda letak kereta	
papillon (de nuit)	moth	con đom đóm; con nhậy (cần quần áo); bướm đêm, sâu bướm	kupu-kupu
Puteri : I've changed as you wanted, but it's the opposite			
nœud papillon	nơ bướm	tali leher	bentuk kupu-kupu

Figure 23 – Extrait de l'entrée vietnamienne du mot « papillon » dans le projet Papillon

un problème au contributeur qui voudrait voir se matérialiser sur le web ses efforts bénévoles. Pour éviter ce délai, les modifications de l'utilisateur lui sont immédiatement visibles dans son *espace personnel*.

Les technologies sous-jacentes sont toutes libres : le serveur est programmé en Java, la base de données est stockée dans un moteur PostgreSQL, et les données sont sous forme XML.

La plate-forme coopérative inclut notamment un service de listes de discussions entre utilisateurs, diligemment archivées. L'effort de recherche est également facilité à l'aide d'un service unifié d'accès à de multiples dictionnaires-ressources et d'un service de partage simple d'information entre utilisateurs, à l'aide de fichiers HTML libre format.

Dans sa thèse [10], Daoud indique que le dictionnaire Papillon héberge 2 millions d'entrées, toutes langues confondues.

Sous-titrage de vidéos. Le sous-titrage de vidéos en ligne les rend accessibles à la fois aux malentendants et aux moteurs de recherche indexant les mots de leur contenu. Ce faisant, les utilisateurs contribuent à la création d'une ressource multimédia son-texte permettant d'entraîner des moteurs de reconnaissance vocale. Amara^{aa} est un exemple de ce genre de plate-forme sans but lucratif, sous-titrant les vidéos des TED Talks, par exemple.

Traduction du site Facebook. Plusieurs sites et logiciels libres s'appuient sur le travail de volontaires pour les traduire et les localiser. Le bénévole acquiert donc une certaine notoriété auprès des utilisateurs du même logiciel.

Les traductions de Facebook sont un excellent exemple, puisqu'une certaine partie du site semble localisée par ses propres abonnés. La Figure 24 illustre comment le géant Facebook s'y est pris pour canaliser les efforts de ses utilisateurs : ils peuvent soit ajouter de nouvelles traductions, soit valider celles produites par d'autres. Des points sont attribués, et un palmarès des meilleurs est disponible. Il n'existe pas de publication sur les données ainsi produites, Facebook étant plus discret que ses abonnés... Une chose est certaine : il y a une partie de validation, et le nombre impressionnant des usagers du site autorise un certain niveau de qualité dans les traductions, probablement par agrégation des résultats (voir

^{aa} <http://www.amara.org/>

Edit Profile

Short translation for Edit Profile

Modifier le profil

+ New translation

{name1} plays this.

list of friends that use/play this app

{name1} joue à cela. (De votre part)

{name1} y joue.

{name1} joue cette.

+ New translation

{name1}, {name2}, {name3}

{name1}, {name2} et {name3}

{name1}, {name2}, {name3}

{name1} {name2} ({name3})

+ New translation

Your Progress

1 translations 4 votes

Your Awards

Translate and vote some more to get awards!

[Check out all the awards!](#)

Translations App Feedback

Click [here](#) for answers to frequently asked questions or to report a bug.

[Translation App Terms of Service](#)

Weekly Leaderboard

[View All](#)

- 1  Sandra Pelletier
- 2  Patrick Perigny
- 3  Gilles Brophy
- 4  Lilith Médusa
- 5  Marc Barbeau
- 6  Ginette Frigon
- 7  Serge T. Gagnon
- 8  Tristan Auclair
- 9  Justin Richard Leung
- 10  Yannick Vézina

Translation Community

[Participate](#)

Figure 24 – Traductions vers le français de Facebook, dans Facebook

section 7.2). L'application Facebook est à mi-chemin entre jeu constructif et application collaborative, empruntant des concepts aux jeux pour faire de la ludification (section 4) d'une tâche « sérieuse ».

Traduction et apprentissage avec Duolingo. Il serait regrettable de mener la présente étude sans parler de l'admirable projet de Von Ahn sur Duolingo [32],[33]. Dans le site web du projet^{bb}, lancé à la fin de 2011, les utilisateurs reçoivent des cours de langue gratuits (5 langues sont offertes) dans une interface d'une très grande qualité où la ludification est reine (Figure 25). L'application en ligne demande à l'apprenant de traduire des segments de phrases, puis des phrases de complexité croissante.

Il est à noter que Duolingo est donc une application coopérative utilisant la ludification, mais n'est pas un jeu constructif, car le but premier de la plate-forme n'est pas récréatif.

^{bb} <http://www.duolingo.com/>

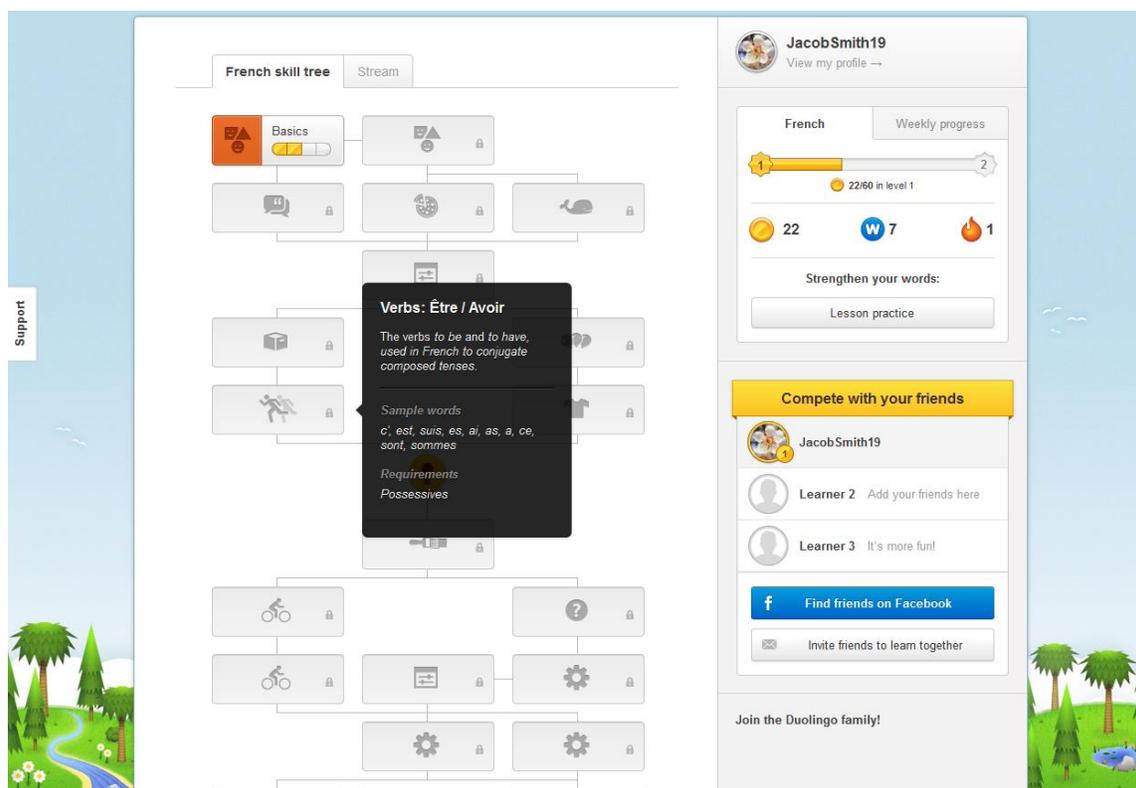


Figure 25 – « Arbre de compétences » pour l'apprentissage du français sur Duolingo

Le véritable travail de coopération, quant à lui, prend la forme d'un mode « immersion », c'est-à-dire tout simplement la traduction d'articles Wikipédia de la langue apprise vers la langue maternelle de l'apprenant. Chaque traduction passe par deux étapes : la production, puis la validation par la communauté. Von Ahn estime que si un million de personnes participaient, tous les articles anglais de Wikipédia pourraient être traduits en espagnol en 80 heures [33], bien que certains condamnent l'amateurisme des traductions produites [34].

8.3 Exemple d'approches coopératives non ludiques rémunérées

Nous ne citerons ici qu'un seul exemple, car l'écrasante majorité des ressources linguistiques créées par ce genre d'approche le sont à l'aide de Mechanical Turk.

Mechanical Turk, d'Amazon. La plus populaire plate-forme de *crowdsourcing* permet de concevoir une myriade de petites tâches (on parle de *microwork*), soumises à des travailleurs, contre rémunération. Cette dernière est très basse (de l'ordre de 0,01 \$ à 0,10 \$ par fragment complété). Ceci permet de construire des ressources linguistiques à très bas prix (un dixième du prix habituel selon [35]), et relativement vite, en soumettant par exemple des tâches d'annotation ou de traduction. Il existe d'autres plates-formes similaires, telles que les sites CrowdFlower (www.crowdflower.com), Clickworker (www.clickworker.com) et Microtask (www.microtask.com).

Un avantage indéniable de Mechanical Turk est la rapidité d'exécution des tâches. Pour des tâches d'annotations simples, on rapporte [9] que les travailleurs les complètent quelques minutes après leur soumission. Dans un cas très simple, une étude [9] avance le chiffre de 500 ms pour accomplir une tâche.

Une personne soumettant une tâche peut spécifier le nombre maximal de tâches qu'un travailleur peut résoudre, des restrictions sur les compétences linguistiques de ce dernier, ou sur sa situation géographique, etc.

Les études divergent cependant sur la qualité des ressources ainsi obtenues. Certaines montrent que la qualité est comparable aux ressources élaborées par des experts lorsque les données sont agrégées (voir section 7.2) et traitées diligemment [9]. D'autres études constatent des défauts significatifs dans la qualité [36], parfois inférieure à celle des outils automatiques [37]. La plate-forme n'est pas à l'abri des travailleurs malhonnêtes, qui expédient leurs tâches afin de gonfler leurs revenus. Passer au peigne fin la qualité des ressources créées avec le service d'Amazon n'est pas le propos de ce rapport.

Dans tous les cas, Mechanical Turk pose des problèmes éthiques importants, puisque les travailleurs n'ont pas de recours en cas d'abus, et n'ont aucune garantie d'un salaire minimum [37]. Ceux qui l'utilisent comme source de revenus principale sont pratiquement des esclaves modernes. Des solutions de rechange équitables comme Samasource^{cc} ont commencé à voir le jour.

Aussi grave est la dépréciation du travail d'annotation. Si les organismes subventionnaires dans la construction de ressources linguistiques voient qu'il existe des solutions très peu onéreuses, quitte à sacrifier la qualité, ne seront-ils pas réticents à mettre l'argent qu'il faut pour créer des ressources de qualité ? Ceci risque à son tour d'éliminer les travailleurs compétents, dont les efforts sont insuffisamment rémunérés, forçant une diminution graduelle de la qualité des annotations.

8.4 Comparaison avec le jeu constructif

On résume ici les différences et similarités les plus notables entre les approches coopératives non ludiques et le jeu constructifs.

8.4.1 Différences

Un avantage notable des approches non ludiques est que leur plus grande polyvalence autorise la création de ressources relativement plus complexes, en particulier dans le cas des approches bénévoles. En général, ces volontaires sont plus à même de bâtir des ressources linguistiques très riches (dictionnaires, encyclopédie, etc.). La simplicité des jeux et plates-formes de microtâches ne conviendrait généralement pas à ces efforts, et les annotations qu'ils produisent sont relativement simples.

Ceci n'est cependant qu'une tendance, et il existe des exceptions.

^{cc} <http://samasource.org/>

On note également pour les approches non ludiques :

- Une **interface plus complexe**, généralement, pour permettre des tâches plus élaborées, où l'utilisateur peut naviguer à sa guise entre un type de sous-tâche et un autre (contribution et validation, par exemple). Dans le cas du *crowdsourcing* dans Mechanical Turk, cependant, l'interface est relativement simple.
- Une **implémentation souvent de type wiki**, c'est-à-dire une application web riche de coopération permettant d'ajouter, modifier et supprimer du contenu, comme pour le projet Papillon et pour le wiktionnaire.
- Des **possibilités de discussion** et de communication sont parfois possibles entre collaborateurs, souvent *via* un wiki, ce qui encourage la mise en commun des talents et l'échange d'idées. Ceci n'est souvent pas possible avec le jeu constructif, où les joueurs sont soit seuls, soit interdits de communication pour éviter la collusion durant une partie.

8.4.2 Similarités

Les approches ludiques et non ludiques ont beaucoup en commun, surtout lorsque ces dernières sont non rémunérées, et qu'il faut alors faire preuve d'ingéniosité pour attirer les bénévoles. On peut relever :

- Des **aspects ludiques** sont présents dans les deux types d'approche, avec scores, badges de récompense, palmarès des meilleurs contributeurs (par exemple traductions Facebook, Duolingo), etc. Dans le cas des approches non ludiques, cela revient donc à en faire de la ludification (section 4).
- La **validation** est souvent présente (par exemple Duolingo, Wikisource (section 3.2), projet Papillon) pour assurer une qualité d'annotation suffisante. En particulier, il est bon de rappeler que les erreurs commises par les annotateurs des jeux constructifs (début de la section 7) se retrouvent également dans ce type d'approche.
- Il faut prévenir l'**abus et la collusion**, et ce, malgré le caractère soit bénévole, soit rémunéré des contributions. Par exemple, dans Mechanical Turk, les travailleurs sans scrupules peuvent expédier leurs tâches. Les créateurs d'ITOLDU [38], qui sert à la construction coopérative d'un lexique bilingue par des étudiants, ont eux aussi rapidement constaté que les étudiants inondaient le système de contributions douteuses pour accumuler des points (non monnayables).

9 Compétitivité financière des approches coopératives

9.1 Dépenses en approches coopératives

À la section 2.3, on a montré les limites des approches traditionnelles, où un ou plusieurs experts élaborent manuellement une ressource linguistique. Avec des coûts avoisinant 1 M USD par million de mots annotés, il serait ruineux d'annoter des volumes de données importants. Les approches collaboratives présentées ici constituent autant de solutions de rechange.

Cependant, il est complexe de les comparer équitablement entre elles. Peu d'études existent sur le sujet, et celles qui chiffrent leurs dépenses tiennent généralement pour acquis que la qualité des annotations produites par les bénévoles est comparable à celle des experts, ce qui est parfois difficilement vérifiable (voir section 7). De plus, les ressources colligées sont très diverses, par conséquent, dégager des généralités est difficile. Enfin, il faut tenir compte de la vitesse à laquelle les ressources sont produites : certains jeux prennent des mois avant de gagner la base d'utilisateurs suffisante pour assurer leur succès ; les plates-formes de microtâches pourraient prendre des années à produire certaines ressources.

Ce qui est clair, c'est que les approches coopératives ne sont pas gratuites, même lorsqu'elles sollicitent le concours de bénévoles ou de joueurs. Il faut notamment budgétiser :

- la préparation des données à annoter ;
- la conception d'une interface d'annotation adaptée ;
- la ludification ou le concept de jeu, le cas échéant ;
- la programmation informatique ;
- le matériel informatique, lorsque les concepteurs se font hébergeurs ;
- les coûts administratifs ;
- les coûts de promotion et de récompenses ;
- le soutien technique ;
- le traitement des ressources produites ;
- la validation des ressources créées, le cas échéant ;
- la lutte contre l'abus éventuel du système (voir section 7.4).

Si l'approche coopérative est rémunérée, il faut naturellement ajouter à cette liste les salaires conséquents.

Un avantage clair des jeux constructifs est le fait que, une fois le jeu implémenté, les annotations ne coûtent plus rien (ou presque), tandis que les approches rémunérées ont un coût fixe par annotation produite.

Nous tentons ici de présenter les mérites économiques de chaque approche. Le lecteur pourra également se référer aux comparaisons financières faites à la fin des sections 5.1 et 5.2.

9.2 Études comparatives [5], [9]

Les analyses comparatives rigoureuses incluses dans les études par Jon Chamberlain, Massimo Poesio et leurs collaborateurs [5], [9] sont fort utiles, ici. Nous les résumons.

Le produit fini d'une tâche d'annotation sont les annotations. Comme on l'a dit plus haut, il faut en agréger plusieurs pour obtenir un accord entre annotateurs (7.2). Notons enfin qu'en moyenne on considère [5] que, dans un texte, un mot sur trois fait l'objet d'annotations. Un corpus d'un million de mots contient donc 330 000 unités à annoter.

Approche traditionnelle entièrement validée. Selon les experts du domaine touchant au financement, elle coûte environ 1 M USD pour annoter un corpus d'un million de mots, et dure plusieurs années. Voir également la section 2.3.

Approche traditionnelle partiellement validée. Elle coûte environ 400 000 USD pour un corpus de 1 million de mots, et dure plusieurs années.

Approche coopérative rémunérée (Amazon Mechanical Turk). En fonction de la difficulté de la tâche d'annotation, elle coûte entre 50 000 USD et 1 M USD pour un corpus d'un million de mots. Ceci inclut entre 5 et 10 répétitions pour chaque annotation, afin d'assurer une certaine qualité. Il faut ajouter environ 5000 USD pour le développement de l'interface d'annotation.

Jeux constructifs Phrase Detective et JeuxDeMots. Le Tableau 3 résume les coûts d'annotation des deux jeux constructifs présentés en section 5.

Tableau 3 – Coûts d'annotation, pour Phrase Detectives et JeuxDeMots, tiré de [9]. Le nombre de mois indique le nombre de mois nécessaire au développement + le nombre de mois de fonctionnement.

	Phrase Detectives (44 mois)	JeuxDeMots (58 mois)
Développement du jeu	40 500 USD	18 000 USD
Publicité et récompenses	9 000 USD	0 USD
Maintenance	34 200 USD	24 300 USD
Total	83 700 USD	41 300 USD
Nombre d'annotations	2,5 M	1,3 M [†]
Coût par annotation	0,033 USD	0,032 USD
Coût par 1 M annotations	33 480 USD	31 770 USD
Coût pour 1 M mots	153 000 USD*	n.d.

[†] Annotations non validées.

* Selon [5], ce qui représente à leur avis la moitié du coût d'Amazon Mechanical Turk.

Le Tableau 3 indique les statistiques d'annotations pour JeuxDeMots après 58 mois de jeu. Après 40 mois de fonctionnement, son créateur principal indiquait [16] que le nombre de parties jouées était de 1,5 M, pour un total de 25 000 h de jeu. En extrapolant à 58 mois de jeu comme dans l'étude de Chamberlain, on obtiendrait 36 250 h de travail. Admettons qu'un (zélé) lexicographe professionnel saisisse 1 relation lexicale (1 annotation) par



Figure 26 – Le jeu OntoPronto. Capture d'écran issue de la vidéo YouTube^{ee}.

minute pour faire le travail d'annotation de JeuxDeMots sur 58 mois. Cela lui prendrait donc 21 700 h de travail environ, soit 650 000 \$ en tout à 30 \$ l'heure.

9.3 Étude de Thaler et coll. [30]

Dans leur étude unique en son genre (à notre connaissance), Thaler et ses collaborateurs [30] ont comparé un jeu constructif appelé OntoPronto avec Amazon Mechanical Turk, pour une même tâche d'annotation. Le jeu, illustré à la Figure 26, est un jeu de sélection à joueur unique ou multiple, de type « accord sur les sorties » (section 6.1.1). La tâche sous-jacente consiste à déterminer si un article de Wikipédia parle d'une entité particulière (par exemple la Suède) plutôt que d'une classe d'entités (par exemple, les pays). Ensuite, il faut catégoriser plus finement l'article dans une hiérarchie de concepts^{dd}.

Ces deux tâches ont également été soumises dans le cadre d'une soumission à Mechanical Turk, d'Amazon (section 8.3), à raison de 0,01 USD pour la première et de 0,05 USD pour la

^{dd} Cette tâche consiste à faire de l'ingénierie d'*ontologie*. Une ontologie est une hiérarchie riche de concepts, servant notamment à l'édification du *web sémantique*, un projet extrêmement prometteur qui permettra à terme des recherches sémantiquement riches et l'intégration de données provenant de sources disparates partout dans le monde.

^{ee} <http://www.youtube.com/watch?v=dek2Sfkgktk>

seconde. L'interface de travail sur Mechanical Turk ressemble au jeu dépouillé de ses aspects ludiques.

Le Tableau 4 montre la comparaison entre les deux approches, incluant le pourcentage d'annotations correctes, après validation manuelle par des experts.

Tableau 4 - OntoPronto versus Mechanical Turk, tiré de [30]

Mesure de comparaison	Jeu OntoPronto	Mechanical Turk
Temps de développement	5 mois-personnes	1 mois-personne
Temps de fonctionnement	1 mois	0,5 mois
Nombre d'annotations consensuelles	2 175	718
Pourcentage annotations correctes	97,4 %	99,2 %
Coût par annotation	s.o. [†]	0,74 USD
Participants	270	16

[†] Le prix par annotation ne s'applique pas. Il n'est pas nul, à cause du coût de développement.

Il est clair que le coût de développement est moindre pour Mechanical Turk, mais au prix d'annotations qui sont payées à la pièce. Il nous semble donc que le nombre d'annotations doit être conséquent pour amortir les coûts de développement plus élevés. Les auteurs relèvent également avec justesse que le nombre limité de participants dans Mechanical Turk pourrait introduire un biais dans les réponses, possiblement gommé dans OntoPronto du fait du grand nombre de joueurs, et donc de points de vue sur une tâche qui a sa part d'ambiguïté (voir le début de la section 7).

Les auteurs concluent en expliquant que, pour la tâche en question, l'approche Mechanical Turk semble supérieure en temps et en argent. Ils mettent également l'accent sur la complexité de mettre au point un jeu constructif enthousiasmant pour la tâche relativement ennuyeuse et parfois complexe qui consiste à classer des concepts.

9.4 Conclusions

La création traditionnelle de ressources linguistiques est très onéreuse (inconcevable au-delà d'un million de mots), mais la qualité produite est très grande. Les approches collaboratives offrent des solutions de rechange très intéressantes. Même si la conception d'un jeu constructif est relativement chère, la perspective de colliger de très grands volumes d'annotations gratuitement est séduisante. Les approches collaboratives non ludiques bénévoles, quant à elles, peuvent s'inscrire elles aussi dans la même configuration avec, peut-être, une motivation scientifique plus significative chez ses participants.

Les approches de *crowdsourcing* à microtâches rémunérées paraissent particulièrement indiquées pour des campagnes d'annotation d'envergure petite ou moyenne, et où la vitesse d'annotation est importante. Cependant, il faut garder à l'esprit qu'une faible diversité d'annotateurs peut appauvrir la ressource, et que l'utilisation de plates-formes comme Mechanical Turk pose nombre de sérieuses questions éthiques.

10 Conclusion et perspectives

En dernière analyse, les approches coopératives, qu'elles soient ludiques ou pas, ont leur place dans la construction de ressources linguistiques. Comme les exemples présentés ici l'ont démontré, des logiciels en ligne, conçus pour canaliser l'*intelligence collective* des internautes avec la convivialité suffisante, peuvent faire émerger des ressources linguistiques de qualité à partir de petites tâches relativement simples.

Les utilisateurs de ces logiciels ne sont pas des experts, mais le volume de leurs contributions est tel que la loi des grands nombres force la convergence de leurs efforts. Cela dit, à notre avis, il reste des tâches pour lesquelles les avis d'experts resteront incontournables. À tout le moins, ceux-ci pourraient s'avérer indispensables pour valider ou pour trancher dans les cas d'ambiguïtés relevées par les utilisateurs d'approches coopératives.

La question du biais introduit par les internautes reste à éclaircir. Sont-ils vraiment représentatifs des locuteurs moyens d'une langue ? Faut-il qu'ils le soient, pour une tâche donnée ? Ces questions sont d'autant plus délicates que certaines applications coopératives posent des questions ouvertes, où l'utilisateur peut répondre librement. Les concepteurs de la plate-forme collaborative doivent se poser ces questions, selon nous.

Nous avons présenté ici des exemples isolés d'application colligeant des ressources. Cependant, il peut être utile d'imaginer ces logiciels si particuliers au sein d'un pipeline de traitement complet. Ainsi, des étapes en amont pourraient pré-colliger et/ou prétraiter une partie des données. Une partie du traitement pourrait faire intervenir un expert, et une autre, l'application coopérative. Enfin, des étapes en aval pourraient accomplir une tâche concrète donnée, peut-être même en temps réel, si le débit des contributions l'y autorise. On peut même concevoir des pipelines combinant plusieurs étapes de coopération différentes, où le contributeur aurait peut-être même la possibilité de « naviguer » d'une étape à l'autre, pour constater le fruit de son labeur en aval, par exemple.

Des stratégies hybrides brouillant les distinctions entre jeux éducatifs, ludification et jeux constructifs pourraient démultiplier le potentiel des approches collaboratives. Dans le domaine de l'enseignement, il y a notamment des efforts considérables investis chaque année en devoirs, exercices et corrections. On pourrait envisager, par exemple, de fournir aux professeurs de français des interfaces de correction des rédactions estudiantines qui leur faciliteraient le travail, et permettraient également de colliger des ressources utiles à la correction grammaticale automatisée. Cette automatisation viendrait à son tour accomplir une partie du travail de l'enseignant, assurant ainsi une certaine motivation de celui-ci. Les étudiants pourraient également faire des exercices de français ou de langue seconde sur des applications collaboratives, qui non seulement enrichiraient leurs connaissances, mais construiraient des ressources variées, un peu comme le fait déjà Duolingo (Figure 25).

Quand les stratégies bénévoles fonctionnent, leur compétitivité dans l'édification de ressources volumineuses est imbattable. Cependant, la nature même de la motivation des contributeurs fait que leur succès n'est pas garanti. Un jeu constructif devra ainsi être adopté avec enthousiasme par une communauté de joueurs notoirement imprévisibles afin d'assurer un certain débit d'annotations. C'est donc dire que le succès des méthodes bénévoles est plus difficilement reproductible. Les plates-formes avec rémunération par microtâche ne sont pas sans défaut : limitées à des tâches restreintes, elles sont trop onéreuses pour créer des ressources importantes et elles pourraient être éthiquement inacceptables, *a fortiori* en milieux universitaire et gouvernemental, réputés plus réceptifs aux inégalités sociales.

11 Références

- [1] T. Chklovski, « LEARNER: A System for Acquiring Commonsense Knowledge by Analogy », in *Proceedings of Second International Conference on Knowledge Capture (K-CAP)*, 2003, p. 4–12.
- [2] J. Giles, « Internet encyclopaedias go head to head », *Nature*, vol. 438, p. 900-901, déc. 2005.
- [3] L. von Ahn et L. Dabbish, « Labeling images with a computer game », in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2004, p. 319–326.
- [4] Nelson Francis and Henry Kucera, *Brown Corpus*. 1964.
- [5] M. Poesio, J. Chamberlain, U. Kruschwitz, L. Robaldo, et L. Ducceschi, « Phrase detectives: Utilizing collective intelligence for internet-scale language resource creation », *Acm Trans Interact Intell Syst*, vol. 3, n° 1, p. 3:1–3:44, avr. 2013.
- [6] S. S. Pradhan, E. Hovy, M. Marcus, M. Palmer, L. Ramshaw, et R. Weischedel, « OntoNotes: A Unified Relational Semantic Representation », in *International Conference on Semantic Computing, 2007. ICSC 2007*, 2007, p. 517-526.
- [7] M. Maugeot-Lerebours, G. Serasset, et M. Lafourcade, « Construction collaborative d’une base lexicale multilingue: Le projet Papillon », *Tal Trait. Autom. Langues*, vol. 44, n° 2, p. 151-176.
- [8] « The Rise of Crowdsourcing by Jeff Howe », *Byliner*. [En ligne]. Disponible sur: <https://www.byliner.com/jeff-howe/stories/the-rise-of-crowdsourcing>. [Consulté le: 21-juin-2013].
- [9] J. Chamberlain, K. Fort, U. Kruschwitz, M. Lafourcade, et M. Poesio, « Using Games to Create Language Resources: Successes and Limitations of the Approach », in *The People’s Web Meets NLP*, I. Gurevych et J. Kim, Éd. Springer Berlin Heidelberg, 2013, p. 3-44.
- [10] M. Daoud, « Utilisation de ressources non conventionnelles et de méthodes contributives pour combler le fossé terminologique entre les langues en développant des “préterminologies” multilingues », Université Joseph-Fourier - Grenoble I, 2010.
- [11] L. von Ahn, B. Maurer, C. McMillen, D. Abraham, et M. Blum, « reCAPTCHA: Human-Based Character Recognition via Web Security Measures », *Science*, vol. 321, n° 5895, p. 1465-1468, déc. 2008.
- [12] L. von Ahn et L. Dabbish, « Designing games with a purpose », *Commun Acm*, vol. 51, n° 8, p. 58–67, août 2008.
- [13] L. von Ahn, « Games with a Purpose », *Computer*, vol. 39, n° 6, p. 92–94, juin 2006.
- [14] L. V. Ahn, R. Liu, et M. Blum, « Peekaboom: A Game for Locating Objects in Images », in *In ACM CHI*, 2006, p. 55–64.
- [15] M. Lafourcade, « Making people play for Lexical Acquisition with the JeuxDeMots prototype », présenté à SNLP’07: 7th International Symposium on Natural Language Processing, 2007.
- [16] M. Lafourcade, « Mémoire d’habilitation à diriger les recherches, Lexique et analyse sémantique de textes : structures, acquisitions, calculs et jeux de mots ». Université Montpellier 2 - LIRMM, 2011.
- [17] G. Sérasset, « Jeux de Mots : à quoi sert un réseau lexical ? », présenté à Fête de la science, Valence, 2009.

- [18] « WordNet », *Wikipedia, the free encyclopedia*. 01-juin-2013.
- [19] J. Chamberlain, M. Poesio, et U. Kruschwitz, « Phrase Detectives - A Web-based Collaborative Annotation Game », in *In Proceedings of I-Semantics*, 2008.
- [20] L. von Ahn, M. Kedia, et M. Blum, « Verbosity: a game for collecting common-sense facts », in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2006, p. 75–78.
- [21] R. Speer, C. Havasi, et H. Surana, « Using Verbosity: Common Sense Data from Games with a Purpose », in *Twenty-Third International FLAIRS Conference*, 2010.
- [22] A. Scharl, M. Sabou, S. Gindl, W. Rafelsberger, et A. Weichselbraun, « Leveraging the Wisdom of the Crowds for the Acquisition of Multilingual Language Resources », in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, Istanbul, Turkey, 2012, p. 379–383.
- [23] D. Vickrey, A. Bronzan, W. Choi, A. Kumar, J. Turner-Maier, A. Wang, et D. Koller, « Online word games for semantic data collection », in *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Stroudsburg, PA, USA, 2008, p. 533–542.
- [24] J. A. L. Farias et F. V. Brasileiro, « Engagement Factors in Games with a Purpose », in *Proceedings of SBGames 2012*, 2012, p. 31–37.
- [25] *1001 Paraphrases: Incenting Responsible Contributions in Collecting Paraphrases from Volunteers*. .
- [26] A. Koller, K. Striegnitz, A. Gargett, D. Byron, J. Cassell, R. Dale, J. Moore, et J. Oberlander, « Report on the second NLG challenge on generating instructions in virtual environments (GIVE-2) », in *Proceedings of the 6th International Natural Language Generation Conference*, Stroudsburg, PA, USA, 2010, p. 243–250.
- [27] N. J. Venhuizen, V. Basile, K. Evang, et J. Bos, « Gamification for word sense labeling », in *Proc. 10th International Conference on Computational Semantics (IWCS-2013)*, 2013, p. 397–403.
- [28] J. Mrozinski, E. Whittaker, et S. Furui, *Collecting a Why-question corpus for development and evaluation of an automatic QA-system*. .
- [29] A. Joubert, M. Lafourcade, D. Schwab, et M. Zock, « Évaluation et consolidation d'un réseau lexical via un outil pour retrouver le mot sur le bout de la langue », arXiv e-print 1201.4279, janv. 2012.
- [30] S. Thaler, E. Simperl, et S. Wölger, « An Experiment in Comparing Human-Computation Techniques », *IEEE Internet Computing*, vol. 16, n° 5, p. 52–58, 2012.
- [31] K. Dahab et A. BELZ, « A game-based approach to transcribing images of text », in *Proceedings of the 7th International Conference on Language Resources and Evaluation*, France, 2010.
- [32] L. von Ahn, « Duolingo: learn a language for free while helping to translate the web », in *Proceedings of the 2013 international conference on Intelligent user interfaces*, New York, NY, USA, 2013, p. 1–2.
- [33] N. Savage, « Gaining wisdom from crowds », *Commun Acm*, vol. 55, n° 3, p. 13–15, mars 2012.
- [34] I. Garcia, « Learning a Language for Free While Translating the Web. Does Duolingo Work? », *Int. J. Engl. Linguist.*, vol. 3, n° 1, p. p19, janv. 2013.
- [35] C. Callison-Burch et M. Dredze, « Creating speech and language data with Amazon's Mechanical Turk », in *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*, Stroudsburg, PA, USA, 2010, p. 1–12.

- [36] A. Kunchukuttan, S. Roy, P. Patel, K. Ladha, S. Gupta, M. M. Khapra, et P. Bhattacharyya, « Experiences in Resource Generation for Machine Translation through Crowdsourcing », in *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC-2012)*, Istanbul, Turkey, 2012, p. 384–391.
- [37] K. Fort, G. Adda, et K. B. Cohen, « Amazon mechanical turk: Gold mine or coal mine? », *Comput Linguist*, vol. 37, n° 2, p. 413–420, juin 2011.
- [38] V. Bellynck, C. Boitet, et J. Kenwright, « ITOLDU, a web service to pool technical lexical terms in a learning environment and contribute to multilingual lexical databases », in *Proceedings of the 6th international conference on Computational Linguistics and Intelligent Text Processing*, Berlin, Heidelberg, 2005, p. 324–332.
- [39] « PIVAX, an online contributive lexical data base for heterogeneous MT systems using a lexical pivot ». [En ligne]. Disponible sur:
http://academia.edu/2819141/PIVAX_an_online_contributive_lexical_data_base_for_heterogeneous_MT_systems_using_a_lexical_pivot. [Consulté le: 09-juill-2013].
- [40] R. Mihalcea, « Open mind word expert: Creating large annotated data collections with web users' help », in *In Proceedings of 4th International Workshop on Linguistically Interpreted Corpora (LINC-03) held in conjunction with EACL*, 2003.

Annexe A : Quelques jeux constructifs en bref

Nota bene. La version électronique de ce document possède des liens entre crochets dans le tableau ci-dessous permettant de naviguer vers les ressources mentionnées. Le texte contient des détails sur la plupart des jeux mentionnés. Les données présentées sont celles obtenues après une durée indiquée dans la colonne « Temps ».

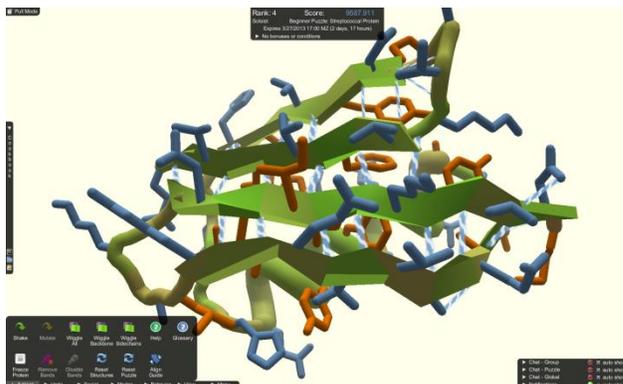
Tableau 5 – Tableau synthétique de quelques jeux constructifs. Les indications entre crochets après les noms de jeux sont respectivement la référence bibliographique, une figure dans ce texte et une url vers le jeu.

Jeu constructif	Joueurs	Temps	Annot.	Qualité
1001 Paraphrases [25] [Figure 20] Collectes de reformulations de phrases pour traductions médicales anglais-perses	< 1300	15 mois	20 944	99,6 % ok
Categorilla, Categodzilla, Free Association [23] [Figure 16] Réseau lexico-sémantique	≈ 2500	n.d.	800 000	moyenne
GIVE games [26] [Figure 13] Évaluation de systèmes de génération de texte	1825	3 mois	n.d.	n.d.
JeuxDeMots, AKI [16], [29] [Figure 8] [url] Réseau lexico-sémantique	2500	40 mois	20 M	75 % ok [†]
Minefield [31] Transcription de texte arabe	n.d.	n.d.	n.d.	n.d.
OntoPronto [30] [Figure 26] Construction d'ontologie	270	1 mois	2 234	97 % ok
Phrase Detectives [5] [Figure 10] [url] Résolution anaphorique	3000	4 ans	2,5 M	≈ experts
Sentiment Quiz [22] [Figure 18] [url] (1) Détection de sentiment dans le texte, ayant servi à l'analyse de l'interprétation de l'information diffusée dans les médias et (2) acquisition de lexique de mots émotifs	3500	11 mois	325 k	≈ experts
Verbosity [20], [21] [Figure 12] [url] Construction d'une base de connaissances générales (<i>common sense knowledge</i>)	30 000	n.d.	500 000	≈ experts
Wordrobe, jeu Senses [27] [Figure 21] [url] Désambiguïsation sémantique	962	1 mois	42 000	≈ experts

[†] 75 % de termes sont suffisamment bien indexés pour s'avérer utiles dans une tâche concrète. Les autres termes ne sont pas nécessairement incorrectement renseignés.

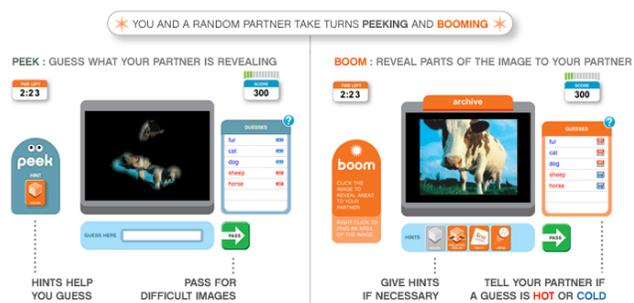
Annexe B : Quelques jeux constructifs non linguistiques

On montre ici très informellement quelques jeux constructifs utilisés à d'autres fins que la construction de ressources linguistiques.



Foldit demande aux joueurs de replier une protéine dans l'espace afin d'en déterminer la conformation 3D la plus plausible.

<http://fold.it/>



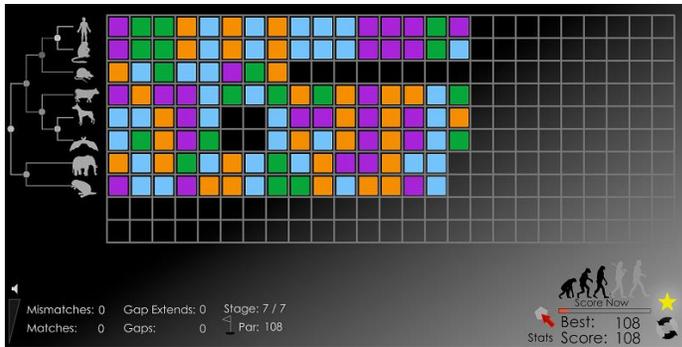
Peekaboom invite un joueur à faire deviner une image à un autre participant en révélant graduellement cette image. [14]

Le jeu permet d'apprendre à un algorithme à localiser des objets dans des images.



Waisda permet l'annotation textuelle d'émissions de télévision en néerlandais.

woordentikkertje.manbijthond.nl



Phylo permet d'aligner des séquences génétiques. Conçu par l'Université McGill.

<http://phylo.cs.mcgill.ca/>



EteRNA propose une structure tridimensionnelle d'acide nucléique et leur demande de trouver une séquence qui est compatible.

<http://eterna.cmu.edu>

Annexe C : Quelques approches coopératives non ludiques

On présente ici une liste non exhaustive d'approches coopératives non ludiques en création de ressources linguistiques. La version électronique de ce rapport contient des hyperliens vers les sites web respectifs.

Tableau 6 – Quelques approches coopératives non ludiques en création de ressources linguistiques

Type de ressource	Plate-forme	Lien
Base de connaissances	LEARNER	[1]
Base de données lexicale pour la traduction	Pivax (dérivé de Papillon)	[39]
Base de données lexicales	Papillon	[7]
Désambiguïsation de sens	Open Mind Word Expert	[40]
Dictionnaire	Wiktionary / Wiktionnaire	site
Interprétation (son)	Babelverse	site
Lexique bilingue créé dans un environnement pédagogique	ITOLDU	[38]
Traduction	Facebook translate	
Traduction	Duolingo	[32], site
Transcription	Ancient lives	site , dans Zooniverse
Transcription	Old weather	site