

Jeux et intelligence collective

résolution de problèmes
et acquisition de données sur le Web

Mathieu Lafourcade
Nathalie Le Brun et Alain Joubert



Version auteurs

Jeux et intelligence collective

Résolution de problèmes et acquisition de ressources
sur le Web
(version auteurs)

par

Mathieu Lafourcade, Nathalie Le Brun, Alain Joubert

Résumé

Les jeux avec but (GWAP en anglais) sont des jeux à travers lesquels l'activité ludique permet de collecter des données ou de résoudre des problèmes trop complexes, ou trop coûteux en termes de moyens humains et matériels pour être résolus par des machines. Ethiquement plus acceptables que bon nombre d'approches par externalisation ouverte (crowdsourcing), ces jeux, qui représentent un type particulier de jeux sérieux, n'en restent pas moins délicats à concevoir, car ils doivent non seulement être attrayants, mais efficaces, c'est-à-dire permettre de récolter des données de bonne qualité, ou encore de progresser significativement dans la résolution d'un problème. Dans cet ouvrage, nous montrons que le concept de GWAP, actuellement en pleine expansion, s'illustre par une grande diversité de formes et d'applications, et se décline dans de multiples domaines, en particulier en biologie, en médecine, et dans le domaine du Traitement Automatique du Langage. Spécialisés dans ce dernier domaine, et nous-mêmes à l'origine, depuis 2007, d'un ensemble de GWAP visant à construire un réseau lexical, (projet *JeuxDeMots*, auquel nous consacrons un chapitre), nous tentons de définir, sur la base de notre enquête sur les GWAP existant, et de notre propre expérience, les critères et conditions qu'il est important de privilégier dans la conception d'un GWAP afin d'en faire un outil d'acquisition de données efficace.

Mots-clés

GWAP, jeux avec but, acquisition de données, résolution de problème, crowdsourcing, externalisation ouverte, TAL, traitement automatique du langage, JeuxDeMots

Table des matières

| | |
|---|-----------|
| Introduction | 1 |
| 1 Jeux biologiques | 7 |
| 1.1 Foldit | 8 |
| 1.2 EteRNA | 11 |
| 1.3 Nanocrafter | 14 |
| 1.4 Phylo | 17 |
| 1.5 Fraxinus | 21 |
| 1.6 Eyewire | 23 |
| 1.7 Citizen sort | 27 |
| 1.7.1 Happy match | 28 |
| 1.7.2 Forgotten Island | 30 |
| 1.8 Le projet Nightjar | 30 |
| 1.8.1 Nightjar game/Nest game | 32 |
| 1.8.2 Egglab game | 32 |
| 2 Jeux à visée médicale | 35 |
| 2.1 Nanodoc | 35 |
| 2.2 Dizeez | 38 |
| 2.3 The Cure | 39 |
| 2.4 Malaria Training Game | 41 |
| 2.5 Malaria Spot Game | 43 |
| 2.6 Worm Watch Lab | 46 |
| 2.7 Play to Cure : Genes in Space | 48 |

| | | |
|----------|---|-----------|
| 3 | Jeux pour le traitement automatique du langage | 51 |
| 3.1 | Pourquoi des ressources lexicales ? | 52 |
| 3.2 | Les GWAP pour le traitement automatique du langage | 53 |
| 3.2.1 | Le problème de l'acquisition des ressources lexicales | 53 |
| 3.2.2 | Ressources lexicales actuellement disponibles | 55 |
| 3.2.3 | Intérêts des GWAP en TAL | 58 |
| 3.3 | PhraseDetectives | 59 |
| 3.4 | PlayCoref | 62 |
| 3.5 | Verbosity | 64 |
| 3.6 | JeuxDeMots | 66 |
| 3.7 | Zombilingo | 67 |
| 3.8 | Infection | 69 |
| 3.9 | Wordrobe | 71 |
| 3.10 | D'autres GWAP dédiés au TAL | 72 |
| 3.10.1 | Open Mind Word Expert | 72 |
| 3.10.2 | 1001 Paraphrases | 73 |
| 3.10.3 | Categorilla/Categodzilla | 73 |
| 3.10.4 | FreeAssociation | 74 |
| 3.10.5 | Entity Discovery | 74 |
| 3.10.6 | PhraTris | 75 |
| 4 | Jeux inclassables | 79 |
| 4.1 | Beat the Bots | 79 |
| 4.2 | Apetopia | 81 |
| 4.3 | Quantum Moves | 82 |
| 4.4 | Duolingo | 83 |
| 4.5 | Le portail ARTigo | 86 |
| 4.5.1 | ARTigo et ARTigo Taboo | 87 |
| 4.5.2 | Combino | 88 |
| 4.5.3 | Karido | 89 |
| 4.6 | Be a martian | 90 |
| 4.7 | Akinator, le génie du Web | 92 |
| 5 | Le projet JeuxDeMots | |
| | – des jeux et des mots | 97 |
| 5.1 | L'objet à construire, un réseau lexical | 97 |
| 5.2 | JeuxDeMots, un jeu associatif | 99 |

| | | |
|-----|---|------------|
| 5.3 | <i>PriClic</i> , un jeu d'attribution | 103 |
| 5.4 | <i>Totaki</i> , un jeu de devinettes | 104 |
| 5.5 | Quelques jeux à votes | 105 |
| 5.6 | Jeux à sélections multiples | 111 |
| 5.7 | Des jeux vers les systèmes contributifs | 115 |
| 5.8 | Données collectées et propriétés des jeux présentés | 118 |
| | Conclusion | 125 |
| | Bibliographie | 133 |
| | Index | 141 |

Version auteurs, dec. 2014

Introduction

Dès le XV^e siècle, l'oxymore *serio ludere* évoque l'idée de combiner les notions de jeu et de chose sérieuse. La littérature humaniste va abondamment user de ce concept pour dénoncer toutes sortes de problèmes de société. Ensuite, jusqu'au développement de l'informatique, ce sont essentiellement les armées du monde entier qui vont exploiter ce concept à travers les *jeux de guerre*, simulations ludiques pour développer de nouvelles tactiques et former les cadres.¹ Le concept moderne du *jeu sérieux* date de 2001-2002, avec le jeu vidéo *America's army*², développé pour l'armée américaine pour simuler des exercices d'entraînement. Mais quasi simultanément, la notion de *serious game* va être popularisée par son application au monde éducatif. Aujourd'hui, la dénomination *jeu sérieux* renvoie à une telle diversité en termes de supports, de concepts, d'intentions, d'approches, et de public visé qu'il est difficile, sinon impossible, d'enfermer cette notion dans une définition exhaustive. L'une des moins restrictives évoque *une application informatique qui combine une intention sérieuse, de type pédagogique, informative, communicationnelle, marketing, idéologique, ou d'entraînement, avec des ressorts ludiques issus du jeu vidéo ou de la simulation informatique*.

Karen Chabriac, en faisant la revue de toutes les tentatives de définition du jeu sérieux³, conclut que la plus synthétique est celle de [Michael et Chen, 2006] : *tout type de jeu dont la finalité est autre que le simple divertissement*.

Du fait de la diversité des thèmes, des objectifs et des approches, il existe de nombreuses tentatives de classification des jeux sérieux, en fonction de leur but (publicitaire, ludo-éducatif, militaire, simulation, prévention, formation, rééducation, etc.), ou encore de la fonction *sérieuse* associée à la base ludique (diffuser un message informatif, éducatif, persuasif, militant, etc., prodiguer un entraî-

¹http://fr.wikipedia.org/wiki/Jeu_s%C3%A9rieux/.

²http://fr.wikipedia.org/wiki/America's_Army.

³<http://www.cndp.fr/savoirscdi/cdi-outil-pedagogique/apprentissage-et-construction-des-savoirs/education-et-pedagogie-reflexion/les-jeux-serieux/les-jeux-serieux-des-elements-de-definition/>.

Introduction

nement visant à améliorer les capacités physiques ou cognitives du joueur, favoriser l'échange de données, entre les joueurs, ou entre le concepteur du jeu et les joueurs), et/ou du segment de marché associé (c'est-à-dire le type de public visé). Il existe même un site interactif dédié à la recherche de jeux sérieux en fonction de divers critères simultanés, répartis en 3 catégories : *Intention*, *Marché*, *Public*⁴.

En réalité, parallèlement à l'idée reçue qu'il n'existe que deux grandes catégories de jeux, les *jeux ludiques* conçus pour se distraire, et les *jeux sérieux* où le joueur s'instruit, il se développe depuis un peu plus d'une décennie une troisième catégorie de jeux, les *Games With A Purpose* ou GWAP, textuellement *jeux avec un but*, où c'est le joueur qui apprend quelque chose à la machine. Le concept de GWAP repose sur l'idée de mettre à contribution les compétences humaines à des fins de recherche et/ou de production de données, que ce soit pour *aider les programmes à progresser dans leur compréhension du monde* ou plus simplement pour utiliser les ordinateurs des particuliers pour accroître la puissance de calcul au service d'un projet de recherche. Dans un cas comme dans l'autre, la composante ludique est essentielle pour motiver le public. Les applications sont multiples et le secteur en pleine expansion. Il s'agit d'exploiter du *temps de cerveau disponible* (disponible et non seulement consentant, mais si possible enthousiaste) pour réaliser des tâches que les machines sont (encore) incapables de faire. Aspect non négligeable, voire capital : ce type de jeux, orienté vers la production de données de toutes sortes, permet d'utiliser ainsi la créativité, l'imagination, les connaissances et le savoir-faire de centaines voire de milliers d'internautes à moindre coût. A noter que contrairement au système AMT⁵ (outil de collecte de ressources *via* le *crowdsourcing* offrant une rémunération dérisoire et des conditions non conformes au droit du travail français) le principe du GWAP ne soulève aucun problème d'éthique, tant qu'il reste gratuit et ne propose pas de lots assimilables à un salaire déguisé [Sagot *et al.*, 2011].

A l'origine de la formalisation du concept de GWAP, il y a les captchas, inventés par Luis Von Ahn [von Ahn, 2006], un universitaire américain : ce sont les petits tests, basés sur le déchiffrement et la resaisie d'une suite de caractères, qui permettent, sur le Web, de différencier un être humain d'un ordinateur, et ainsi d'empêcher le spam, le phishing ou toute autre activité malveillante par des moyens *automatiques*. Prenant conscience que les 10 secondes passées par un être humain à déchiffrer un captcha (donc à faire quelque chose qu'un ordinateur ne sait pas faire) pouvaient être utilement employées, Luis Von Ahn crée *Recaptcha* : désormais, lorsqu'on déchiffre un captcha, non seulement on s'identifie comme être humain, mais on aide à numériser des livres, en déchiffrant des suites de caractères que la reconnaissance optique de caractères (OCR) ne parvient pas à déchiffrer. Le principe du GWAP est né, et va être illustré par *ESP Game* [von Ahn et Dabbish, 2004] : le père des captchas invente un jeu consistant à présenter la même image à deux joueurs qui vont gagner

⁴<http://serious.gameclassification.com/FR/>.

⁵Amazon Mechanical Turk.

Introduction

des points et progresser dès lors qu'ils proposent les mêmes mots-clés pour la définir. L'intérêt est naturellement de rendre plus performante, précise, rapide et pertinente la recherche d'images par mots-clés dans un moteur de recherche.

Le recours à des GWAP, que ce soit pour collecter des données, faire annoter des images ou documents, ou solliciter le public pour la résolution de grands problèmes scientifiques, est actuellement en pleine expansion et concerne tous les domaines, même si c'est en sciences de la vie et en médecine qu'il donne lieu aux expériences et aux résultats les plus médiatisés. Le concept de *science citoyenne* témoigne non seulement d'un changement dans la manière dont les questions scientifiques sont perçues par le public, mais aussi d'une volonté de faire sortir la science des laboratoires et les chercheurs de leurs tours d'ivoire. En rendant les enjeux de la recherche accessibles et compréhensibles par le commun des mortels, on démystifie et on désacralise la science. Pour les chercheurs, l'approche ludo-collaborative est un puissant moyen d'impliquer le public et de susciter son adhésion et son empathie, tout en sollicitant et en valorisant son regard extérieur de non-spécialiste. Signalons que, selon Luis Von Ahn, dans les années 2005-2010, près de 10^{10} (soit 10 milliards) heures étaient dépensées annuellement par les individus jouant sur Internet (ce qui, à l'échelle planétaire, correspond à une moyenne légèrement supérieure à 1 heure par individu et par an). Pourquoi ne pas essayer d'en détourner, ne serait-ce qu'une infime partie, pour des jeux utiles permettant l'acquisition de ressources ou la résolution de problèmes ?

Les GWAP foisonnent donc dans de multiples disciplines, et force est de constater, après un tour d'horizon minutieux, que l'effort consenti pour donner à une tâche utile une dimension réellement ludique est très inégal. En effet, assez rares sont les jeux où l'intérêt est davantage stimulé par un véritable challenge, une émulation entre joueurs, un système de classement motivant, et surtout un réel intérêt de la tâche sous-jacente, plutôt que par l'idée d'aider la science et/ou de faire une bonne action. Fréquemment, divers paramètres sont empruntés à l'univers du jeu (design, avatar, sons, ...) pour donner un vernis ludique à une tâche monotone et répétitive. Il est clair que dans de nombreux cas, les concepteurs du jeu misent avant tout sur la *fibres citoyenne* des joueurs, à travers l'engouement suscité par la science dite *participative* et le sentiment extrêmement valorisant de faire quelque chose d'utile dans un domaine qui reste prestigieux aux yeux du grand public.

En matière de GWAP, et bien que tous reposent sur le principe du *crowdsourcing*, on note de grandes disparités selon la nature et l'ampleur de la tâche, le public visé, le domaine de recherche, les compétences requises chez les joueurs, la dimension ludique et/ou pédagogique, la manière dont les données produites par le jeu sont traitées. Le choix de critères pour établir une classification est donc malaisé, mais on peut néanmoins juger intéressante la classification de [Good et Su, 2013] qui repose sur la nature, l'ampleur et la complexité de la tâche. Ces auteurs distinguent en effet, dans les systèmes exploitant le *crowdsourcing* à des fins scientifiques, les *microtâches* qui sont des tâches

Introduction

pouvant être résolues en quelques secondes par toute personne capable de lire quelques instructions simples, et les *macrotâches* qui concernent au contraire des problèmes complexes, qui résistent aux experts qualifiés de la recherche institutionnelle. Les premières nécessitent un grand nombre de personnes qui vont traiter un énorme volume de données en peu de temps, et dont les contributions, (fortement) redondantes entre les joueurs, vont être agrégées pour fournir des données d'aussi bonne qualité que des annotations d'experts. Pour les secondes, le recours au *crowdsourcing* permet de détecter, au sein d'une large population de candidats potentiels aux compétences très hétérogènes, les quelques personnes de talent, qui grâce à l'environnement interactif fourni, vont non seulement manifester mais développer l'inventivité, la curiosité et la créativité nécessaires pour relever le défi et permettre de réelles avancées scientifiques.

Notons que parmi les GWAP, une énorme majorité concerne des *microtâches*.

Notons également que les jeux à forte audience, dont les résultats spectaculaires se prêtent à une large couverture médiatique, et qui sont en général des *macrotâches* (*Foldit*⁶, *EyeWire*⁷) ont tous un site internet dédié, alors que les GWAP d'ambition plus modeste, relevant plutôt des *microtâches*, sont souvent proposés *via* des portails. Le plus connu et le plus ancien est *Zooniverse*⁸.

A l'origine du portail *Zooniverse*, il y a *GalaxyZoo*⁹, un projet astronomique en ligne qui symbolise un peu les débuts de la science dite *citoyenne*. *GalaxyZoo* est un projet scientifique basé sur le volontariat, destiné à caractériser les galaxies à partir de photos, et le succès fut tel (en 2007, 85 000 contributeurs en 5 mois, d'après Wikipedia¹⁰) que l'idée de s'adjoindre la collaboration du grand public a inspiré de nombreux autres programmes. Actuellement le portail *Zooniverse* est une énorme plate-forme internationale, qui fédère un grand nombre de projets collaboratifs parmi lesquels certains se sont parés des atours du jeu pour avoir l'air plus séduisants. Après s'être inscrit (plus d'un million de personnes inscrites dans le monde), il est possible de collaborer à tous les projets du portail, qui sont classés par grands thèmes (espace, météorologie, nature, humanités, biologie). Au sein de chaque thème, diverses activités de *crowdsourcing* sont proposées, quasiment toutes dévolues au traitement d'une masse volumineuse de données, que ce soit pour identifier faune et flore sur des photos de fonds marins (*Seafloor Explorer*), ou des animaux sauvages qui passent dans le champ d'une caméra installée dans une réserve naturelle (*Snapshot Serengeti*), ou encore déchiffrer les étiquettes manuscrites de spécimens animaux ou botaniques conservés dans des muséums d'histoire naturelle (*Notes from Nature*), et bien d'autres tâches fort utiles et pour lesquelles un traitement automatisé est exclu. Toutes ces activités relèvent davantage de la participation bénévole que du jeu,

⁶<http://fold.it/portal/>.

⁷<http://eyewire.org>.

⁸<https://www.zooniverse.org>.

⁹<http://www.galaxyzoo.org>.

¹⁰http://fr.wikipedia.org/wiki/Galaxy_Zoo.

Introduction

malgré la présence d'éléments ludiques tels qu'un *score*, ou la constitution d'une *collection* réunissant tous les spécimens traités, ou encore l'octroi de *grades* sanctionnant l'importance du travail effectué (passer de *cadet* à *lieutenant* puis *capitaine* en déchiffrant et en transcrivant les informations météorologiques de journaux de bord de navires américains du milieu du XIX^e siècle dans *Old Weather*, pour contribuer à l'étude de l'évolution des climats). C'est pourquoi, à l'exception de *Worm Watch Lab*, nous n'avons pas décrit ces jeux, pour lesquels le côté ludique ne s'apparente qu'à une vague surcouche assez peu convaincante, la motivation étant plutôt stimulée par l'idée valorisante de participer à la *vraie science* (*Real Science Online*).

En résumé, nous pouvons dire que les GWAP sont des jeux, oui mais des jeux utiles ; ce sont des jeux utiles, oui mais des jeux. Ils sont utiles pour la communauté, ce sont des jeux pour les joueurs. Les concepteurs de GWAP ne doivent pas perdre de vue cette dualité, sous peine soit d'obtenir une ressource de très faible volume (si les joueurs se lassent trop vite), soit d'obtenir une ressource de faible qualité (si l'aspect ludique est développé au détriment de l'aspect utilitaire). Nous reviendrons plus en détail au chapitre 5, ainsi que dans la conclusion de ce livre, sur les caractéristiques et limites des GWAP. Les GWAP sont donc des jeux à finalité, même si tous les jeux ont une finalité, majoritairement se distraire ou apprendre. On devrait plutôt parler de *jeu à finalité d'acquisition de ressources* ou dans certains cas de *jeu à finalité de résolution de problèmes*.

Dans cet ouvrage nous présentons et analysons un certain nombre de GWAP selon une organisation thématique. Bien entendu, cette liste est loin d'être exhaustive : le temps de la rédaction et de l'impression de cet ouvrage, de nouveaux GWAP auront apparus (et certains auront disparus). Le chapitre 1 se focalise sur les GWAP issus du domaine de la biologie au sens large. Le chapitre 2 s'intéresse plus précisément aux jeux ayant une visée médicale (mais pas ceux à visée thérapeutique, qui ne sont pas des GWAP). Le chapitre 3 décrit des GWAP concernant la langue, la linguistique et le traitement automatique du langage. Le chapitre 4 est celui des GWAP ne rentrant dans aucune des catégories ci-dessus. Le chapitre 5 présente et analyse en détail le projet *JeuxDeMots* dont l'objectif est de construire une grande base de connaissances lexicales à l'aide de jeux. Enfin, la conclusion tire quelques leçons de l'expérience *JeuxDeMots*, et du tour d'horizon des autres projets, pour tenter de définir quels critères il est important de privilégier dans la conception d'un GWAP afin d'en faire un outil efficace, pour l'acquisition de données ou la résolution de problèmes complexes.

Introduction

Version auteurs, dec. 2014

CHAPITRE 1

Jeux biologiques

Dans les sections suivantes, nous présentons les principaux GWAP actuellement disponibles en biologie et biochimie.

Dans le domaine très spécialisé de la biologie moléculaire, les résultats de *Foldit*¹ ont été tellement positifs (trois articles publiés dans la revue *Nature* – [Cooper *et al.*, 2010], [Marshall, 2012] et [Hand, 2010] –, et un autre dans les *Actes de l'Académie des Sciences* [Khatib *et al.*, 2011]) et le retentissement médiatique tellement important (un grand nombre d'articles dans la presse généraliste et de vulgarisation scientifique) que le concept du GWAP s'est rapidement propagé en biologie. De plus en plus de chercheurs tentent, avec plus ou moins de bonheur, de recourir à une approche ludo-contributive pour résoudre des problèmes complexes.

Zoran Popovic, un des créateurs du jeu, et directeur du Center for Game Science de l'Université de Washington, estime que son laboratoire est à l'origine de l'émergence d'une nouvelle voie de découverte scientifique, qui tire parti des capacités de créativité d'une large population, dès lors que l'on sait l'impliquer dans les mécanismes de recherche *via* une initiation efficace.

Adam Gazzaley, professeur de neurologie, physiologie, psychiatrie, à l'Université de Californie et créateur de *Neuroracer* (un jeu pour améliorer les capacités cognitives des gens âgés), plus mesuré, pense que le domaine du jeu sérieux est encore à ses débuts et qu'il faut trier entre ce qui prétend être basé sur la science et ce qui a été réellement validé par des méthodes scientifiques. Se-

¹*Foldit* est accessible à l'adresse <http://fold.it/portal/>.

Jeux biologiques

lon lui, il faut multiplier les approches et les outils et vérifier les résultats avant de pouvoir faire des recommandations fortes.¹

1.1. Foldit

Type de tâche concernée : macro-tâche

Remarque préliminaire : *Foldit*¹ est certainement l'ancêtre des GWAP biologiques dévolus à la résolution d'un problème complexe. Un de ses concepteurs, Adrien Treuille, a également participé à la création de *EteRNA*² (voir section 1.2), et l'équipe de *Foldit* est également à l'origine du *Center for Game Science*, structure de l'Université de Washington consacrée à l'univers du jeu à but scientifique. Le Center for Game Science héberge un tout nouveau GWAP biologique : *Nanocrafter*² (voir section 1.3), un jeu de construction de nanostructures moléculaires.

Lancement : mai 2008

Audience/popularité : 57 000 joueurs 2 ans après le lancement.

But : réaliser des modèles prédictifs de structures tridimensionnelles de protéines à partir de leur composition en acides aminés.

Intérêt du crowdsourcing pour le problème à résoudre : prévoir la structure tridimensionnelle d'une protéine en fonction de sa composition en acides aminés est un enjeu fondamental : un grand nombre de maladies sont causées par des mutations qui, en affectant la structure tridimensionnelle d'une protéine, altèrent son fonctionnement normal. Si on peut prévoir la structure *normale* et stable d'une protéine, on peut comprendre comment une mutation se traduit au niveau de la conformation spatiale, et développer des thérapies adaptées (vaccins, protéines à visée thérapeutique. . .).

A l'origine de *Foldit* il y a le programme de calcul distribué Rosetta@home² : il s'agit d'un programme qui détermine les conformations possibles que peut adopter une protéine en calculant les probabilités d'interactions entre les segments de chaînes polypeptidiques en fonction des acides aminés qui les composent, les interactions étant d'autant plus probables que le niveau d'énergie requis pour qu'elles aient lieu est minimal. Ainsi, la structure tridimensionnelle la plus probable est celle qui nécessitera le moins d'énergie pour se constituer. Cependant, pour une séquence d'acides aminés donnée, il existe tant de conformations spatiales possibles que la simulation *via* des algorithmes est extrêmement longue et complexe. Rosetta@home utilise les ordinateurs des particuliers et s'exécute quand ceux-ci sont peu ou pas utilisés, accroissant ainsi la puissance de calcul au service du projet dit *distribué*. *Foldit* a été conçu comme le prolongement interactif et ludique de Rosetta@home.

D'après David Baker [Hand, 2010], un des chercheurs à l'origine du projet, on ne demande pas aux joueurs de *Foldit* une simple compétence de reconnaissance visuelle, de classification, on

²*Nanocrafter* est accessible à l'adresse <http://nanocrafter.org/>.

Jeux biologiques

leur demande de mettre leur logique, leur intuition, leur capacité à voir dans l'espace, leur esprit d'initiative et leur créativité au service de la résolution de problèmes complexes d'optimisation. Baker dit utiliser trois compétences du cerveau humain, trois talents typiquement et exclusivement humains :

- la capacité de savoir prendre un risque à court terme pour un avantage à long terme et l'inverse ;
- la capacité à voir dans l'espace ;
- la capacité de se rendre compte qu'on est dans une *impasse* et de s'arrêter à temps (c'est-à-dire avant d'avoir gaspillé trop de temps).

Zoran Popovic, un des principaux acteurs du projet (dans son volet informatique), affirme que pour le moment, le cerveau humain reste meilleur que la machine dès que le paramètre *créativité* devient essentiel et incontournable. De plus, la dimension de jeu crée les conditions idéales pour l'émergence d'une expertise collective : les concepteurs de *Foldit* font le constat que non seulement les joueurs s'améliorent au fil du temps, mais aussi qu'ils s'organisent pour additionner leurs compétences respectives. Enfin, le jeu est conçu de telle sorte que les trouvailles des joueurs sont réinjectées, recyclées pour être intégrées dans les mécanismes de manipulation de la protéine virtuelle, de telle sorte que les nouveaux joueurs deviennent experts de plus en plus rapidement.

Ainsi, un cas typique (illustrant la première des trois capacités évoquées ci-dessus) où l'humain est plus performant que la machine, est celui où une structure doit être profondément remaniée pour optimiser sa configuration, c'est-à-dire arriver à une forme stable avec une dépense énergétique minimale. Pour la machine, cela implique trop de sacrifices en termes de temps de calcul, puisque cela nécessite, dans un premier temps, de désordonner un agencement existant.

L'habileté des joueurs les plus passionnés à manipuler les chaînes protéiques s'est révélée telle que, rapidement, les concepteurs ont fait évoluer le jeu de manière à offrir la possibilité de créer des protéines *ex-nihilo* (applications potentielles : trouver des molécules susceptibles de catalyser certaines fonctions métaboliques, fabriquer des candidats vaccins, concevoir des molécules satisfaisant à des contraintes biologiques précises pour des applications biotechnologiques ou médicales...).

Par ailleurs David Baker utilise les trouvailles des joueurs, c'est-à-dire des stratégies de pliage particulières, pour les automatiser afin d'améliorer les algorithmes de prédiction de structures protéiques. La communauté de joueurs est considérée comme un partenaire scientifique à part entière, et associée comme coauteur aux publications scientifiques [Khatib *et al.*, 2011].

Qualités/compétences humaines sollicitées : intuition, logique, vision dans l'espace (« *Humans have 3D understanding that computers just can't handle yet, and computers have the number crunching capabilities that humans can't possibly compete with.* »³), esprit de déduction, esprit d'initiative, curiosité, créativité.

Apprentissage nécessaire/accessibilité du jeu : une étape d'initiation, nécessaire pour se familiariser avec le concept et les outils, est dispensée *via* un tutoriel. Il consiste en une succession de 31

Jeux biologiques

puzzles qui dévoilent par des explications pas-à-pas tous les mécanismes et modes d'actions disponibles pour agir sur la structure tridimensionnelle proposée afin de l'amener dans une conformation précise. On peut se confronter aux vrais *puzzles* et entrer en compétition avec les autres joueurs avant d'avoir résolu les 31 étapes du tutoriel, mais la relative complexité des mécanismes à maîtriser en fonction des divers cas de figure possibles rend la chose assez aléatoire. La manipulation de la structure modélisée en 3D s'avère assez ardue et déroutante, et il arrive fréquemment que la résolution d'un *puzzle* doive plus à un heureux hasard qu'à une réelle maîtrise des outils proposés. On est pleinement dans la configuration de *hardgame* décrite par [Good et Su, 2013].

Education du joueur : lors de la phase d'apprentissage *via* le tutoriel, le joueur apprend les règles biochimiques fondamentales qui régissent le repliement des chaînes protéiques : il doit minimiser les espaces vides au sein de la molécule en favorisant l'établissement de ponts hydrogène, manipuler les chaînes latérales de telle sorte que les parties hydrophobes soient orientées vers l'intérieur et éviter une trop grande proximité entre les atomes. La modélisation de la molécule permet de visualiser ces paramètres et les mécanismes qui permettent d'agir dessus. Sans devenir des biochimistes experts, les joueurs réguliers et passionnés finissent par acquérir une grande expertise vis-à-vis de la conformation spatiale des protéines, et la faculté de certains à comprendre intuitivement ce qu'il faut modifier pour améliorer la structure en 3D justifie aux yeux des concepteurs l'intérêt de confier à des non-spécialistes le design de protéines virtuelles, potentiellement synthétisables.

Interactivité : elle est très développée : les joueurs n'interagissent pas à travers le jeu proprement dit, puisque chacun est seul avec sa structure protéique à optimiser, mais les possibilités de communiquer avec les autres joueurs et avec les concepteurs sont multiples : une fenêtre de messagerie instantanée est présente dans l'interface où l'on manipule la protéine, et on voit également le nombre de joueurs connectés en même temps sur la même configuration (plusieurs *puzzles* sont proposés simultanément). Depuis la page d'accueil du jeu, où les dernières nouvelles sont actualisées quasi quotidiennement, on peut accéder à un blog, où les concepteurs détaillent les avancées scientifiques et les résultats obtenus grâce à l'activité des joueurs, ainsi qu'à un forum et à un wiki.

Challenge/émulation : à l'issue de la phase d'apprentissage, où l'on s'exerce seul et à son rythme sur des modèles de protéines dont la structure est connue, on s'attaque à des *puzzles*, donc des structures protéiques à optimiser, qui sont mis en jeu pour une durée déterminée. On peut choisir soit un *puzzle* niveau débutant, ou un *puzzle* en fonction de sa ou ses caractéristiques (*overall* (généraliste), *prediction* (composition en acides aminés connue, structure 3D inconnue), *symmetry* (protéines constituées de plusieurs éléments identiques disposés suivant un axe de symétrie). On est alors en compétition en temps réel avec tous les autres joueurs qui ont choisi ce *puzzle* et qui s'évertuent à en améliorer la structure. Il y a entre 0 et une cinquantaine de joueurs connectés simultanément sur le même *puzzle*, et les choses prennent un tour assez excitant puisqu'il s'agit de faire

Jeux biologiques

monter son score et ainsi de s'élever dans le classement, que l'on peut suivre dans une petite fenêtre incrustée dans la page. Le compteur de points réagit à la moindre intervention sur la structure, et une animation assortie d'un bruitage apparaît sur la page dès que l'on a dépassé un autre joueur (*rank up*), mais aussi si on régresse dans le classement. Le but recherché est évidemment de stimuler le joueur pour qu'il « trafique » sa protéine jusqu'à arriver en tête du classement. La limitation dans le temps pimente le défi.

Design/ambiance : la modélisation en 3D de la molécule est très réussie, on peut la manipuler et l'orienter aisément à la souris et ainsi la voir facilement sous de multiples angles afin de repérer les endroits problématiques (vide trop important, atomes trop proches, liaison hydrogène à favoriser. . .). L'ensemble est très réactif, chaque action produit un effet, heureusement toute action est annulable, et la possibilité de revenir à la meilleure configuration obtenue est rassurante et autorise toutes les initiatives : on retrouve son meilleur score, et la conformation correspondante, même si on se lance dans une tentative d'amélioration qui finalement empire les choses (ce qui est fréquent. . .). Le bruitage des différentes actions est agréable, la musique de fond n'est pas lancinante, et de toutes façons, l'interface est paramétrable avec notamment la possibilité de désactiver les sons.

1.2. EteRNA

Type de tâche concernée : macrotâche

Lancement : fin 2010

Audience/popularité : plus de 50 000 inscrits, beaucoup d'articles dans la presse généraliste et scientifique, dont un dans les *Actes de l'Académie des Sciences*, où les joueurs sont associés comme coauteurs [Lee *et al.*, 2014].

But : le but de *EteRNA*³ est de concevoir virtuellement des molécules d'ARN complexes qui seront synthétisées réellement, et testées pour vérifier qu'elles adoptent bien la conformation tridimensionnelle théoriquement prévue, et qu'elles sont actives biologiquement. Étant donné le rôle-clé que joue l'ARN dans la synthèse des protéines, et le rôle que jouent les protéines au sein du vivant, on comprend que la maîtrise de la synthèse et de la conformation de l'ARN est à la base de multiples applications médicales, thérapeutiques, et biotechnologiques.

Intérêt du crowdsourcing pour le problème à résoudre: les cerveaux humains s'avèrent plus efficaces que les ordinateurs pour prédire les manières dont une molécule de séquence nucléotidique connue peut se replier. Il est à noter que cette assertion semble controversée : la NP-difficulté ne semble pas prouvée pour le design d'ARN, comme le rapporte [Tchernia, 2014] (un problème est dit

³*EteRNA* est accessible à l'adresse <http://eterna.cmu.edu/web/>.

Jeux biologiques

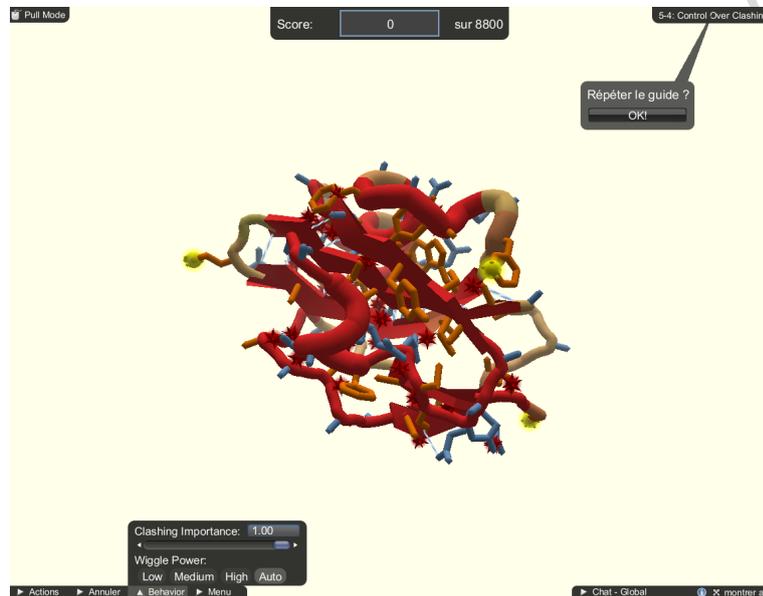


Figure 1.1.* Foldit. L'interface de jeu est une vue tridimensionnelle de la protéine étudiée. On manipule les éléments moléculaires figurés (ponts hydrogène, chaînes hydrophobes, etc.) avec la souris, et chaque action modifie la conformation spatiale de la protéine. Le symbole * indique qu'une version couleur de la figure est disponible à l'adresse : <http://www.iste.co.uk/fr/lafourcade.pdf>

NP-difficile quand il ne peut pas être résolu en un temps raisonnable par une méthode algorithmique et qu'il justifie de ce fait le recours à l'approche participative).

Qualités/compétences humaines sollicitées : logique, capacité de déduction, initiative, curiosité, créativité, stratégie...

Apprentissage nécessaire : le joueur est initié aux règles d'appariement des bases nucléotidiques *via* un tutoriel consistant en une série de *puzzles* qu'il peut résoudre grâce aux explications données étape par étape.

Education du joueur : bien que les règles d'appariement des bases soient très simples, les paramètres qui déterminent la stabilité de la molécule ne peuvent être appréhendés que progressivement, au fur et à mesure des configurations spatiales rencontrées, et en lisant ce que publient les joueurs assidus, qui finissent par acquérir une véritable expertise dans ce domaine hyperspécialisé de la biologie moléculaire.

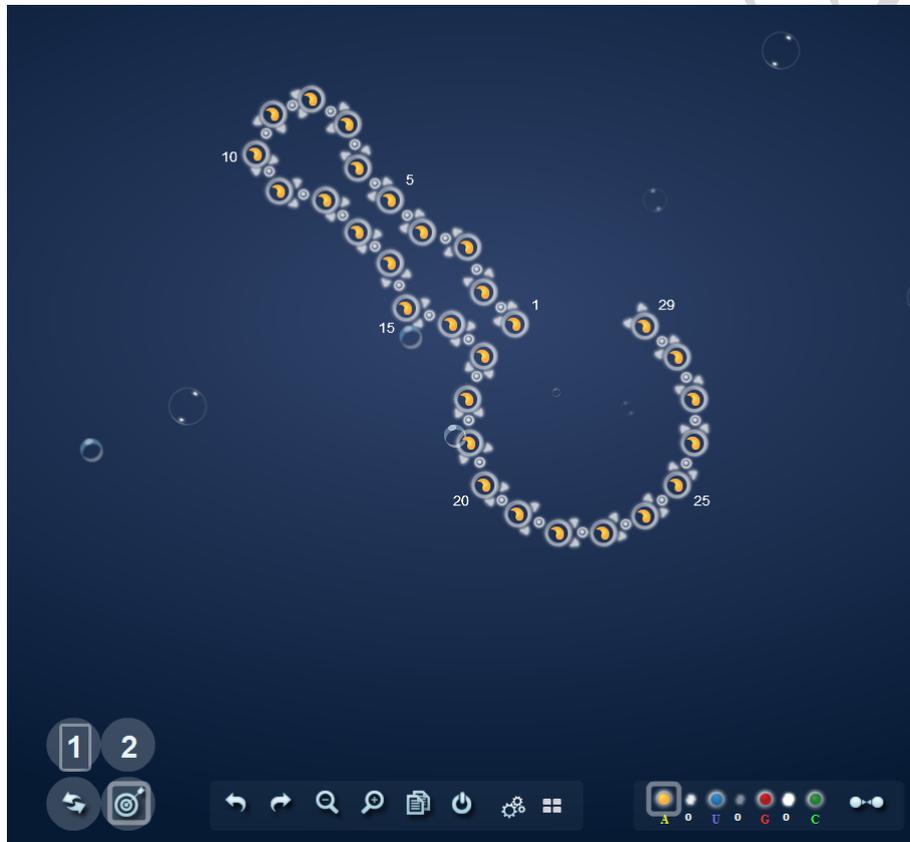


Figure 1.2.* EteRNA. Ici une molécule d'ARN dans la conformation spatiale qu'elle doit adopter si on réalise l'enchaînement de nucléotides approprié, celui qui va permettre l'établissement de liaisons garantissant la stabilité de cette structure. Les formes jaunes sont à remplacer par les bases A, T, G, C que l'on choisit en bas à droite.

Interactivité : comme pour *Foldit*, l'interactivité entre les joueurs est favorisée : la messagerie instantanée du jeu est très active et montre quasiment toujours entre 40 et 50 joueurs connectés simultanément, lesquels semblent former une communauté d'habitues qui se connaissent et s'interpellent. D'autre part, pour chaque *puzzle* proposé, on a accès à un espace de commentaires où les joueurs qui le souhaitent peuvent s'exprimer, donner des indices, ou expliquer comment ils en sont venus à bout. Enfin, le forum d'*EteRNA* est également très fréquenté et on note aussi un wiki, essentiellement alimenté par la communauté de joueurs, qui regorge de stratégies, de conseils, et d'explications pour réussir les *puzzles*.

Jeux biologiques

Challenge/émulation/ludification de la tâche : les joueurs ne sont pas en concurrence directe les uns avec les autres, ils résolvent des *puzzles* pour gagner des points, monter dans le classement et franchir des *grades* multiples : *Tutorial completed*, puis *Puzzle master LV1, LV2*, etc. (en fonction du nombre de puzzles résolus) puis *Top Player LV1, LV2*, etc. (faire partie des n meilleurs joueurs). Parallèlement, en résolvant des *puzzles* on accumule des *points*, on devient successivement un *Lab Member* puis un *Puzzle Architect LV1, LV2, LV3*. Le grade ultime est celui de *Millionaire*, réservé aux joueurs ayant accumulé 1 000 000 de points en jouant à *EteRNA*. D'autre part, la convivialité et les interactions entre joueurs sont également des paramètres du classement, et sont vivement encouragées par des *grades* : *comment* (avoir laissé au moins 5 commentaires), *Follow players* (suivre au moins 5 joueurs), *gather followers* (avoir au moins 5 *followers*). Remarque : la multitude de *grades* et de paramètres de classement : nombre de puzzles réalisés, rang dans le classement, points gagnés, récompenses de convivialité, font que le joueur, même débutant, voit immédiatement des *compteurs* progresser (points, nombre de *puzzles*, rang, ...), ce qui est une forme de stimulation et d'encouragement. Son profil lui montre ce qu'il a *débloqué*, mais aussi toutes les étapes qu'il lui reste à parcourir.

Design/ambiance : le design est soigné, l'interface claire, l'aide bien présente et accessible, les couleurs harmonieuses... Des animations et sons rythment chaque action, l'ergonomie est agréable, on peut zoomer à volonté sur la molécule. Le bleu profond éclairé d'un halo et les bulles qui montent lentement donnent l'illusion d'un milieu aquatique, reposant, serein, zen.

1.3. Nanocrafter

Type de tâche concernée : macrotâche

Remarque : *Nanocrafter*² a été lancé fin avril 2014 par la même équipe que *Foldit*, dans le cadre du Center for Game Science de l'Université de Washington, avec ces mots de Zoran Popovic :

« *After more than a year in the making, NanoCrafter is ready for the first public Beta release. Over the many years to come, we expect the game to evolve almost on a daily basis towards a vibrant community of creative DNA designers, hopefully producing many exciting novel designs. As with any Beta we expect many things need to be improved, we look forward to improving the game on a daily basis. I am announcing the beta release of the game at Games for Change in NYC today.* »²

But : *Nanocrafter* aborde le thème de la synthèse de fragments d'ADN en utilisant les propriétés d'appariement des bases formant l'ADN, et notamment les mécanismes de déplacement de brins qui surviennent spontanément quand plusieurs simples brins de séquences compatibles sont mis en présence. Il s'agit, à travers le jeu, de réfléchir sur la construction de nanostructures dont les

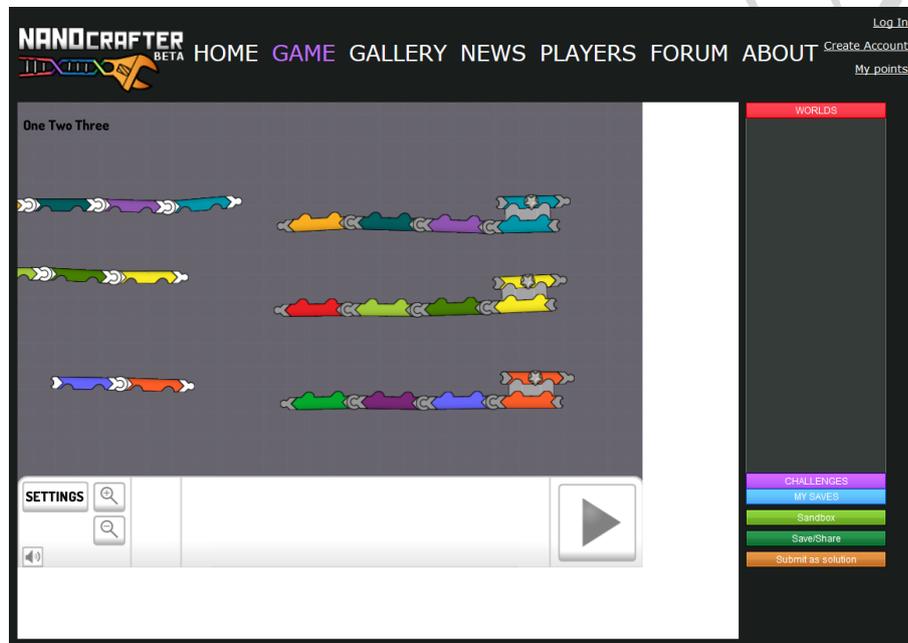


Figure 1.3.* Nanocrafter. Dans la configuration ci-dessus, les éléments colorés représentent des fragments d'ADN simple brin qu'il faut assembler en respectant les correspondances de couleur pour former des structures double-brin

applications biotechnologiques potentielles suscitent un intérêt croissant en biologie structurale et notamment en nanomédecine.

Intérêt du crowdsourcing pour le problème à résoudre: *Nanocrafter* est directement inspiré de ce qui fonctionne dans *Foldit* et *EteRNA*, ce sont donc les mêmes compétences humaines qui sont recherchées à travers les tâches proposées. La principale différence est l'accent mis sur la créativité, puisqu'il s'agit d'inventer des assemblages moléculaires en vue de créer des structures qui n'existent pas dans la nature, en partant des briques constitutives du vivant (l'ADN) et de ce que l'on sait de leurs contraintes et propriétés d'assemblage. Le domaine de la biologie synthétique est encore balbutiant, mais les applications en matière de santé sont potentiellement considérables, et, toujours d'après Zoran Popovic, c'est le domaine privilégié pour que s'exprime la longueur d'avance que le cerveau humain possède encore sur la machine.⁴

Qualité/compétences humaines recherchées : créativité, curiosité, esprit d'initiative, logique.

Apprentissage nécessaire/accessibilité du jeu : comme dans les autres jeux de ce type, le joueur débutant est initié aux mécanismes qui régissent l'assemblage des nucléotides, et doit franchir les

Jeux biologiques

différentes étapes de l'initiation en effectuant des tâches de construction simples au début, et qui se complexifient progressivement. Le joueur évolue à travers des *mondes* qui correspondent à des techniques d'assemblage moléculaire qu'il faut maîtriser, chaque monde étant subdivisé en plusieurs seuils qu'il faut franchir pour débloquent le suivant. Il y a donc un minimum de paramètres biochimiques à connaître et savoir exploiter pour se lancer dans la biosynthèse de nanostructures biologiques.

Education du joueur : comme dans *Foldit* et *EteRNA*, même si le joueur ne connaît rien à la biochimie au départ, il acquiert, par le simple fait de jouer, les notions basiques d'appariement des bases nucléotidiques, et le minimum de règles à savoir pour manipuler la molécule, règles qui deviennent cependant de plus en plus compliquées au fur et à mesure que l'on franchit des niveaux. Comme dans les GWAP précédents, il peut s'en contenter, et n'a pas besoin d'en savoir plus pour jouer. Le jeu est encore trop récent, mais on imagine volontiers que, comme pour *EteRNA* et *Foldit*, les joueurs les plus *accros* vont rapidement acquérir une expertise nourrie par leur pratique d'une part, l'émulation d'autre part, et la fréquentation assidue des espaces de discussion du jeu (forum et news), où l'équipe de conception, très présente, communique régulièrement des informations et évolutions, avec un effort de vulgarisation plus ou moins marqué. On a l'impression que ce sont les non-scientifiques qui, en s'appropriant les codes, le vocabulaire, et en discutant des problématiques entre eux acquièrent progressivement une culture très spécialisée et centrée sur les problèmes soulevés par le jeu, les chercheurs, de leur côté, faisant le minimum de vulgarisation pour que le grand public comprenne ce qu'il faut faire (le jeu dispose d'une rubrique d'aide très détaillée). Le rapprochement entre les chercheurs et le grand public se fait davantage dans le sens où le grand public acquiert des connaissances de spécialistes, que dans celui où les spécialistes se mettent à la portée du public...

Interactivité : le jeu vient d'être lancé, mais comme pour ses aînés, les possibilités d'interaction et de communication sont très développées : les joueurs disposent d'une fenêtre de messagerie instantanée incrustée dans la fenêtre de jeu, d'un forum, et un wiki viendra probablement compléter le tout.

Challenge/émulation/ludification : le jeu est trop récent pour en juger, mais comme on ne change pas une équipe qui gagne, les mécanismes qui ont fait le succès de *Foldit* sont plus ou moins repris. En marge d'une progression solitaire à travers des *mondes* où on débloquent progressivement les niveaux en découvrant de nouvelles manières d'agir sur les éléments de construction, des *challenges* sont proposés, qui sont des objectifs à remplir avec des contraintes à respecter, comme par exemple *réalisez un fragment d'ADN double brin, en utilisant les pièces ci-dessous*. Quand on estime qu'on a atteint l'objectif, on soumet la solution trouvée, qui rapportera des points si elle est validée. Pour le moment, rien n'explique clairement comment et par qui les solutions proposées sont validées, ni

Jeux biologiques

le mode de calcul des points, sans doute parce que le jeu débute. Donc, assez peu d'émulation et de mise en concurrence, (on ne voit pas si plusieurs joueurs tentent simultanément le même challenge, on ne sait pas si les joueurs votent pour les différents designs proposés), l'essentiel de l'intérêt vient du côté *casse-tête*, et du suspense lié à la validation ou pas de la structure proposée. A noter que dans les niveaux des différents *mondes*, la réponse (valide ou pas) est donnée dès la soumission, avec en prime, le signalement des assemblages problématiques en cas d'échec. Inversement, dans les *challenges*, on n'a pas de nouvelles immédiates de la solution que l'on propose.

Design/ambiance : l'interface est agréable, claire, colorée, les éléments de construction se déplacent à la souris avec fluidité, la manipulation est intuitive, les règles d'assemblage sont accessibles à tout moment *via* un bouton dans la fenêtre de manipulation, on peut mettre une structure à la poubelle et recommencer, en cas d'échec on a le choix entre repartir à zéro ou rappeler la dernière configuration réalisée et la modifier pour tenter de corriger ce qui ne va pas.

1.4. Phylo

Type de tâche concernée : macrotâche

Généralités : le créateur et son équipe mettent l'accent sur le fait que *Phylo*⁴ est un projet libre et définissent leur jeu comme un nouvel outil d'aide à l'analyse du génome à la disposition de la communauté scientifique internationale. Un lien *contribute to Phylo* invite à collaborer en améliorant le code à la base du jeu, en traduisant le jeu dans différentes langues (l'ambition affichée est que *Phylo* soit traduit dans toutes les langues possibles ; il l'est dans 11 actuellement). De plus, depuis octobre 2013, les chercheurs du monde entier peuvent télécharger leurs alignements de séquences sur *Phylo* et les soumettre à la sagacité des joueurs pour les améliorer. Cela se fait *via* une interface baptisée *Open-Phylo*, qui nécessite une inscription distincte, et à travers laquelle les chercheurs peuvent non seulement télécharger leurs propres séquences, qui doivent avoir été préalablement, mais aussi générer et organiser les *puzzles* sur les portions de leur choix [Kwak *et al.*, 2013].

Jérôme Waldispühl, en mettant en avant le statut *open-source* de son projet, soulève le problème du statut des données générées par *crowdsourcing*, en déplorant que la plupart du temps elles ne soient pas libres. Il précise également que les premiers projets de *crowdsourcing* ont, du fait de la nouveauté du concept, bénéficié d'une couverture médiatique considérable qui a énormément facilité leur diffusion. Le soutien par des médias forts, à très large audience, est la condition *sine qua non* pour que ce type de projet rencontre le public qu'il mérite. Or, le nombre de projets dits de *science citoyenne* est en augmentation constante, l'effet de nouveauté va donc en s'atténuant, et la banalisation du concept va rendre de plus en plus difficile et aléatoire l'adhésion du public. Pire encore,

⁴*Phylo* est accessible à l'adresse <http://phylo.cs.mcgill.ca/>.

Jeux biologiques

selon [Hand, 2010], journaliste à la revue *Nature*, si le concept de GWAP s'étend démesurément, il arrivera un moment où le grand public ne sera plus du tout émerveillé ni flatté d'être sollicité pour faire avancer la science. De là à dire que les gens risqueront au contraire de se sentir exploités et manipulés, et que le jour viendra où les scientifiques devront payer des volontaires pour s'assurer le concours du public, il y a un pas qui est encore loin d'être franchi, mais Michael Kearns, informaticien à l'Université de Pennsylvanie appréhende que ne se développe une activité économique autour du concept de *science citoyenne* [Hand, 2010].

Pour le créateur de *Phylo*, les portails de type *Zooniverse*⁵, *Citizen Science Center*⁶, ou *Scistarter*⁷ peuvent dans une certaine mesure compenser le problème du manque de couverture médiatique auquel vont se heurter les nouveaux projets au fur et à mesure que le concept de GWAP va se banaliser. Pour ne pas diluer la puissance potentielle de la foule dans une multitude de projets, il suggère que les équipes de recherche se fédèrent et utilisent la même interface de jeu quand ils ont le même type de problème à résoudre, plutôt que de créer chacun leur petit GWAP dans leur coin. C'est pourquoi il met *Phylo* à la disposition de la communauté scientifique internationale. C'est ce qu'il explique dans le blog du jeu, en présentant l'interface *Open-Phylo*. L'existence d'un jeu comme *Fraxinus*⁸ (voir section 1.5), plus ou moins calqué sur *Phylo*, est le contre-exemple parfait de ce que préconise Waldispühl.

Lancement/auteur : lancé en Octobre 2010 par Jérôme Waldispühl et son équipe de l'Université McGill à Montréal, le jeu a été largement commenté dans les médias scientifiques [Strauss, 2012] et généralistes, et comme *Foldit* et *EteRNA*, il a fait l'objet d'une publication scientifique à laquelle les joueurs ont été associés en tant que coauteurs [Kawrykow *et al.*, 2012].

Audience : plus de 300 000 personnes depuis le lancement d'après le média scientifique en ligne *Biomedical Computation Review*⁵.

But : les chercheurs (biochimistes, généticiens, évolutionnistes) ont besoin de comparer les séquences de macromolécules (notamment ADN, mais aussi ARN, protéines...) issues de différentes espèces, afin de détecter les zones similaires. Une des applications est de déduire du taux de similarité entre les séquences des informations précieuses sur les relations de parenté entre les espèces, mais aussi d'identifier, grâce aux zones d'alignement, des sites fonctionnels importants, ou encore mettre en évidence des mutations, et comprendre l'origine de maladies génétiques. Les alignements de séquences multiples sont réalisés par des algorithmes, et représentent typiquement un problème *NP-difficile* (c'est-à-dire ne pouvant être résolu en un temps raisonnable par un ordinateur étant donné la complexité et le coût en termes de durée de calcul, sans même évoquer le coût financier),

⁵*Zooniverse* est accessible à l'adresse <https://www.zooniverse.org/>.

⁶*Citizen Science Center* est accessible à l'adresse <http://www.citizensciencecenter.com/>.

⁷*Scistarter* est accessible à l'adresse <http://scistarter.com/>.

⁸*Fraxinus* est accessible à l'adresse <https://apps.facebook.com/fraxinusgame/>.

Jeux biologiques

ce qui justifie l'approche ludo-contributive. Le but est donc de faire améliorer par des joueurs les alignements de séquences obtenus *via* les algorithmes de calcul. Le jeu se présente comme un casse-tête où il s'agit de trouver le meilleur alignement vertical entre des successions de blocs de couleurs qui représentent les 4 bases de l'ADN. *Phylo* permet donc de faire faire par le grand public ce que faisaient généralement les chercheurs manuellement : réorganiser les morceaux de séquences pour tenter d'améliorer l'alignement fourni par les logiciels.

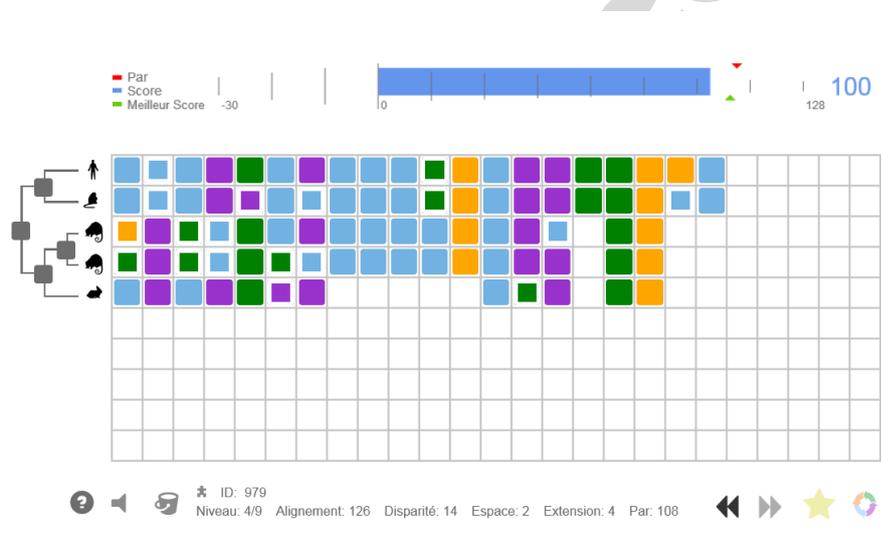


Figure 1.4.* Phylo. Il s'agit, en déplaçant latéralement les carrés de couleur, de trouver le meilleur compromis d'alignement possible entre les 5 fragments de séquences présentés. Au-dessus des séquences, la barre bleue illustre la progression du score, le petit triangle rouge étant le score réalisé par l'ordinateur, qu'il faut atteindre et dépasser pour passer au niveau suivant.

Intérêt du crowdsourcing pour le problème à résoudre : trouver le meilleur compromis d'alignement pour un grand nombre de séquences est donc un problème *NP-difficile*. En réalisant qu'effectuer cette tâche manuellement ne nécessitait aucune connaissance en génétique mais simplement un bon sens de l'observation, et des capacités de raisonnement et de logique, Jérôme Waldispühl a eu l'idée de traduire le problème d'alignement de séquences multiples en un jeu de manipulation de formes géométriques colorées, et a fait le constat que dans plus de 40 % des cas, les joueurs sont capables d'améliorer l'alignement issu des algorithmes de calcul. « Contrairement à un algorithme, l'humain est capable de contourner la règle du jeu pour arriver à l'alignement »⁶ affirme-t-il.

Qualités/Compétences humaines recherchées : sens de l'observation et de la déduction, raisonnement, logique.

Jeux biologiques

Apprentissage nécessaire/accessibilité du jeu : *Phylo* a été conçu pour être aussi simple et rapidement accessible que le célèbre Tetris. La simplicité de la tâche à effectuer est telle qu'aucun apprentissage n'est nécessaire, et que le jeu est immédiatement accessible : tout le monde, à tout âge, est capable de faire glisser des carrés de couleurs le long de lignes superposées, en cherchant à créer des colonnes de couleur identique et comme les séquences (lignes) ne sont pas identiques, il faut trouver le meilleur compromis entre disparités de couleurs inévitables et *trous* dans la ligne. Un tutoriel est néanmoins proposé, mais essentiellement pour expliquer comment sont calculés les points et quelles stratégies doivent être privilégiées pour améliorer le score.

Education du joueur : contrairement aux jeux comme *Foldit*, *EteRNA* ou *Nanocrafter*, où le joueur s'initie aux règles d'appariement des bases, acquiert des connaissances sur la solidité comparée des liaisons atomiques, doit tenir compte de notions telles que la concentration, et connaître et utiliser diverses propriétés biochimiques, *Phylo* ne requiert aucune connaissance génétique ou biochimique et le jeu proprement dit ne dispense aucune information d'ordre scientifique. Il s'agit là d'un parti-pris visant à réduire le problème à résoudre à un jeu aussi simple et populaire que possible, un casse-tête coloré qui ressemble à n'importe quel petit jeu en ligne où l'on peut faire rapidement une partie pour se détendre. La science devient presque invisible... A l'accueil du jeu, le sous-titre *un puzzle pour contribuer à la recherche sur les maladies génétiques* ainsi que la possibilité de choisir un *puzzle associé à une maladie* sont néanmoins là pour inciter à jouer en titillant la corde sensible de la bonne conscience. Naturellement, le joueur curieux peut facilement trouver à quoi sert *Phylo*, et l'information scientifique est disponible *via* un blog et des *news* en page d'accueil du jeu⁴, qui sont essentiellement des liens vers des articles de presse. Comme le souligne Jérôme Waldispühl, « Le joueur se distrait mais ce n'est pas du temps perdu, et à la fin de sa partie on lui dit sur quel gène relié à quelle pathologie il a permis d'avancer ».¹⁰

Interactivité : elle n'est quasiment pas développée. *Phylo* est un jeu solitaire, comme les *réussites*... On essaie de faire mieux que l'ordinateur, et la seule référence aux autres joueurs est le classement. Il existe un forum, mais il n'est pas du tout actif. A l'accueil, une fenêtre montre les tweets relatifs au jeu en temps réel.

Challenge/émulation : l'intérêt principal du jeu est sa prise en main immédiate et l'assurance de marquer très facilement des points dans un premier temps, suffisamment pour donner envie au joueur de *s'accrocher* quand ça devient plus difficile. Ainsi, un *puzzle* est composé de plusieurs *niveaux*, chaque niveau ajoutant des séquences à aligner. Égaler le score de l'ordinateur (le *Par*) suffit pour accéder au niveau suivant, et c'est plutôt facile dans les premiers niveaux. Lorsque le joueur parvient à terminer le *puzzle*, il accède aux statistiques et informations sur ce *puzzle*, qui sont un élément de motivation, puisqu'il y lit son score, le score moyen, mais aussi le meilleur score obtenu pour ce *puzzle* et le nom du détenteur du *record*... Il a alors la possibilité de rejouer le *puzzle*, pour tenter

Jeux biologiques

d'égaliser ou surpasser le *record*. Trois classements permettent d'apprécier sa progression : mensuel, hebdomadaire et global ; le classement hebdomadaire est très motivant parce qu'on y progresse vite. Réaliser un *record* est également flatteur et incite à rejouer. L'objectif de Jérôme Waldispühl était de créer un jeu de type Tetris, auquel n'importe quelle personne puisse s'adonner quand elle a un peu de temps, et cela semble totalement réussi : les règles sont suffisamment simples pour qu'une partie suffise à les assimiler, et une partie est suffisamment rapide pour que l'investissement temporel ne soit pas un frein.

Design/ambiance : le design du jeu est agréable, l'interface personnalisable (on peut modifier la couleur du plateau et des bases constitutives de l'ADN, activer/désactiver les sons des actions et la musique de fond), on peut choisir la langue. La manipulation des blocs est fluide, l'ergonomie satisfaisante, on voit immédiatement le résultat de ses actions grâce à une barre de progression bien visible, ainsi que le score à battre (le *Par*). On regrette néanmoins que l'alignement réalisé par l'ordinateur ne soit pas dévoilé lorsqu'on ne parvient pas à égaler le *Par* et qu'on décide d'abandonner. Cela serait instructif...

1.5. Fraxinus

Type de tâche concernée : macrotâche

But/lancement : ce jeu⁸ est un jeu d'alignement de séquences en tous points comparable à *Phylo*, lancé en décembre 2012 par le Laboratoire de Sainsbury, un laboratoire de recherche anglais spécialisé dans les pathologies végétales. A la différence de *Phylo*, il est centré exclusivement sur la recherche concernant la maladie du frêne causée par le champignon *Chalara* et n'est accessible qu'aux membres du réseau social Facebook.

Les joueurs, en améliorant des alignements de séquences issues de l'agent pathogène et de l'arbre victime fournissent des informations susceptibles d'aider les chercheurs à comprendre comment le champignon tue les arbres et surtout comment et pourquoi certains frênes sont résistants, ce qui est d'une importance capitale pour élaborer une stratégie de lutte et éradiquer la maladie qui décime les frênes du Royaume-Uni.

Intérêt du crowdsourcing pour le problème à résoudre/Qualités humaines requises/Accessibilité du jeu : exactement les mêmes que pour *Phylo*.

Education du joueur : comme pour *Phylo*, et pour les mêmes raisons, il n'est pas nécessaire d'avoir des connaissances en génétique, ou en biochimie pour être efficace dans *Fraxinus*, mais une information assez détaillée sur la maladie, les ravages causés sur les arbres, et les stratégies de recherche pour éradiquer ce fléau est dispensée au début du jeu, et sert à susciter l'adhésion et la motivation en jouant sur les cordes sensibles du patriotisme et de l'engouement pour la sauvegarde

Jeux biologiques

de la nature : « Aidez-nous, sinon 90 à 95 % des frênes du Royaume Uni vont mourir... » A la fin de chaque *puzzle* réussi apparaît la mention « *You are helping to save the ash tree* » en guise de piqûre de rappel...



Figure 1.5.* Fraxinus : des séquences à aligner. Il s'agit, en déplaçant latéralement les séquences au moyen des flèches gauche et droite, de trouver la meilleure manière de faire coïncider verticalement les couleurs des blocs constituant les 6 séquences, entre elles, ET avec la séquence de référence visible en haut.

Interactivité : le jeu, accessible uniquement *via* Facebook profite naturellement des facilités du réseau social pour tenter de développer l'interactivité entre les joueurs et surtout les inciter à populariser le jeu auprès de leurs *amis*. On est vivement incité à inviter des *amis* à jouer, à comparer son score avec ceux de ses *amis*, à partager ses résultats, et bien sûr à parler du jeu sur son *mur*.

Challenge/émulation : lorsqu'on se lance dans un *puzzle* d'alignement, le *record* à battre apparaît dans la fenêtre de jeu, avec le nom de son détenteur. Il suffit de l'égaliser pour s'en emparer, et on est alors invité à défier ses *amis* de faire un meilleur score sur le même *puzzle*. On peut soit se lancer dans un nouveau *puzzle*, soit tenter de « voler » un *record* détenu par quelqu'un d'autre, en améliorant l'alignement. Tous les *puzzles* réalisés apparaissent dans un volet du jeu sous forme d'icônes, mentionnant le score obtenu et portant la mention *improve* ou *try again* suivant que l'on a battu le *record* ou non. Il est stimulant de détenir un *record* et manifestement le jeu essaie de valoriser cet aspect et de susciter l'envie de prouver soit qu'on peut faire mieux, soit qu'on est le meilleur.

Jeux biologiques

De plus, les concepteurs affirment que les noms des détenteurs de records seront mentionnés dans les banques de données d'alignement de séquences et dans les publications.

Design/ambiance : l'interface est agréable, claire, bien agencée, la manipulation des séquences intuitive, la barre de progression stimulante, le record à battre est bien en évidence, l'aide accessible. Le design est réussi, les 4 bases nucléotidiques sont représentées par des petites feuilles de 4 couleurs harmonieuses.



Figure 1.6.* Fraxinus : des séquences presque alignées. Par rapport à la figure précédente, on constate que les blocs de couleurs des 6 séquences sont presque parfaitement alignés, à la fois entre eux et avec le modèle. Mais bien que le score soit passé de -148 à $+425$, l'alignement peut encore être amélioré puisqu'on peut lire, en bas à gauche, que le record à battre est 520.

1.6. Eyewire

Type de tâche concernée : macrotâche

But/Lancement : Eyewire⁹, développé et lancé en décembre 2012 par l'équipe de Sebastian Seung, neurobiologiste au Massachusetts Institute of Technology, est en relation avec le projet du connectome humain, financé par l'institut national de la santé américain. Le connectome est un projet colossal qui vise à établir la cartographie des relations neuronales dans et entre toutes les aires corti-

⁹Eyewire est accessible à l'adresse <http://eyewire.org/>.

Jeux biologiques

cales du cerveau humain. *Eyewire* est un jeu conçu pour aider à la reconstitution des prolongements des neurones en 3D à partir de photographies de coupes histologiques sériées de rétine. A partir de la cartographie tridimensionnelle des neurones, les chercheurs en neurosciences vont pouvoir identifier les synapses et mettre en relation les connexions et l'activité cérébrale. Seung et son équipe viennent de publier dans *Nature*, avec *The EyeWriters* en coauteurs, des résultats importants sur la perception visuelle du mouvement obtenus grâce aux données fournies par le jeu [Kim *et al.*, 2014]. Cette avancée valide, selon eux, le recours au *crowdsourcing* via le jeu en ligne pour faire progresser rapidement et efficacement les grandes questions scientifiques.

Audience : en mai 2014 : Plus de 135 000 joueurs recrutés et 150 pays représentés (source : Wikipedia ¹⁰).

Intérêt du crowdsourcing pour le problème à résoudre : l'œil et l'intelligence humaine combinés se révèlent bien plus performants que les programmes informatiques pour détecter les limites et contours cellulaires, en particulier lorsque que les prolongations des fibres nerveuses s'entrelacent en un réseau dense, comme c'est le cas dans la rétine. Le joueur est confronté à un *cube* qui représente un morceau de rétine traversé de part en part par une portion de branche neuronale qui a été partiellement et grossièrement délimitée automatiquement par analyse d'image. Il s'agit pour le joueur d'améliorer le traçage, en *coloriant* ce qu'il identifie comme appartenant à ce neurone sur une photo de microscopie électronique qui représente le même *cube* en coupe transversale. Chaque *cube* est découpé en 256 tranches. Les touches fléchées du clavier permettent de se déplacer dans l'épaisseur du *cube* en passant d'une tranche à l'autre et de *suivre* la branche neuronale en corrigeant le travail de l'algorithme d'analyse, c'est-à-dire en *coloriant* les portions de neurone éventuellement omises, voire en détectant des ramifications également ignorées par l'ordinateur. Un neurone entier correspond environ à 1000 *cubes*, chacun de ces 1000 *cubes* est analysé par 3 à 5 joueurs différents, et c'est la concordance entre leurs tracés qui va légitimer la cartographie finale.

Compétences humaines requises : les concepteurs insistent sur le fait que pour devenir un *Eyewriter* performant, nulle connaissance en neurologie, ni même en biologie générale n'est requise. Le jeu est accessible à toute personne dotée d'un bon sens de l'observation et de la conception dans l'espace.

Education du joueur : même si aucune connaissance en neurologie n'est nécessaire, les joueurs disposent d'une profusion d'informations scientifiques sur le projet auquel ils participent en jouant, et sont régulièrement tenus au courant non seulement des évolutions du jeu, mais des avancées scientifiques qui en résultent. D'autre part, les joueurs les plus anciens et expérimentés se voient confier le pouvoir de superviser le travail de la communauté. On leur reconnaît donc une véritable expertise dans la capacité visuelle à détecter le tracé d'un neurone sur une photo de microscopie

¹⁰<http://en.wikipedia.org/wiki/EyeWire>.

Jeux biologiques

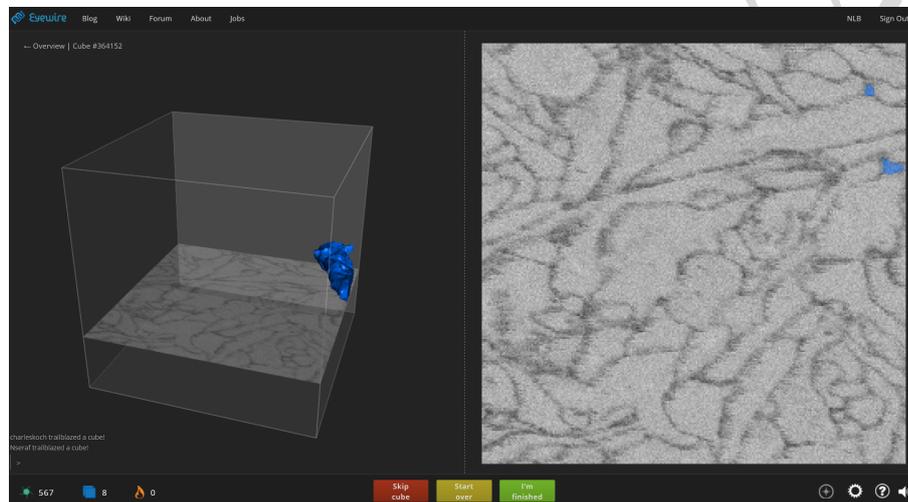


Figure 1.7.* *L'interface d'Eyewire. A gauche le « cube » (fragment de coupe histologique de rétine) qui contient la branche neuronale à délimiter, dont on aperçoit un fragment en bleu. A droite la coupe transversale correspondante dans laquelle le joueur « colorie » les zones qui lui semblent appartenir au neurone. Il « suit » ainsi la branche neuronale en parcourant toute l'épaisseur du cube, à travers laquelle il se déplace avec les touches fléchées du clavier.*

électronique, donc un document sur lequel le commun des mortels ne voit qu'une mosaïque dans des camaïeux de gris.

Interactivité : bien que la tâche soit plutôt monotone et que chaque joueur colorie son petit bout de neurone dans son coin, la convivialité et l'interactivité semblent très développées dans le jeu. Une fenêtre de messagerie instantanée est disponible dans l'interface, et les joueurs présents discutent abondamment et semblent former une large communauté d'habitues qui s'interpellent et plaisantent, mais viennent également au secours de tout nouveau venu qui pose une question. D'autre part, la fenêtre de messagerie rend également compte de l'activité du jeu et donne une idée du nombre de joueurs connectés en communiquant en temps réel les points obtenus par ces derniers. Une manière de susciter l'émulation et une occasion, pour eux, de se congratuler.

Challenge/émulation : étant donné la relative aridité de la tâche (répétons-le, il s'agit de scruter des photos grisâtres à la recherche de continuité et d'homogénéité), on sent une volonté de valoriser et d'inciter les joueurs à rivaliser en les récompensant *via* un système de points, de titres et de classement assez complexe. Le joueur reçoit d'autant plus de points que son tracé est proche de la moyenne des tracés des autres joueurs, et il touche également une *prime* au prorata du temps passé, censée récompenser son application et le soin apporté à son *coloriage*. A travail égal, certaines cel-

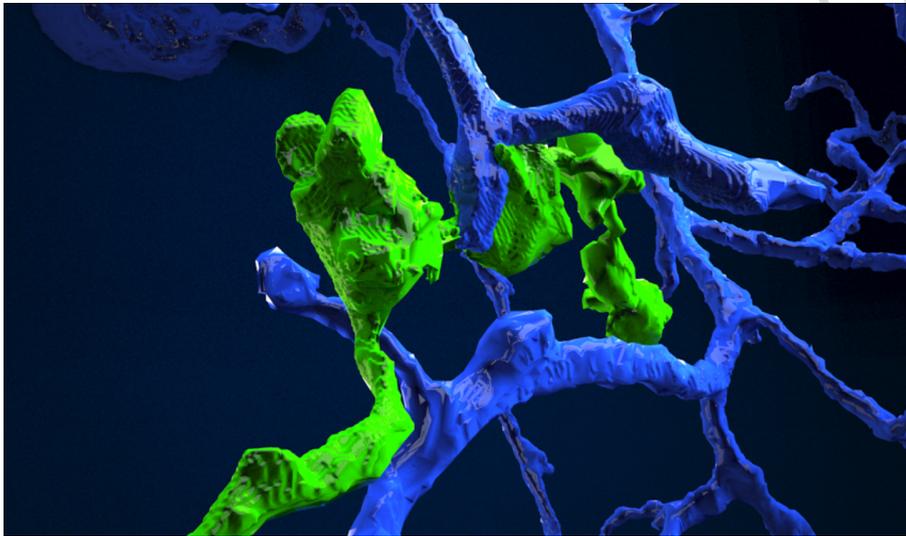


Figure 1.8.* Eyewire. Cette image issue du blog du jeu met en évidence une synapse (le point de jonction entre deux neurones) découverte grâce à l'activité des joueurs

lules rapportent le double car elles sont réputées plus difficiles à isoler (*starbust cells*). Parallèlement aux points, la méticulosité du joueur et la pertinence de ses tracés sont évaluées par un pourcentage de précision (*accuracy*). Dès que le joueur a atteint 80 % de précision il est autorisé à être *pionnier*, c'est-à-dire à être le premier à intervenir sur un *cube*, et bénéficie ainsi d'une *prime de défrichage* et gagne à nouveau des points chaque fois qu'un autre joueur traite ce *cube*. Les joueurs qui ont fait leurs preuves et satisfait à des tests peuvent être promus aux grades de *Scout* ou *Scythe*, ce qui leur donne le droit de superviser le travail des autres joueurs et de valider ou invalider certains tracés. Ainsi, la complexité du classement, la diversité des paramètres évalués, l'ancienneté, ainsi que la difficulté des cellules traitées créent une hiérarchie au sein des joueurs et entretiennent l'émulation. De plus, les compétitions organisées sont un moyen de fidéliser les joueurs, qui ont l'occasion de rivaliser sur un objectif précis. Ces *challenges* sont de différentes natures : par exemple, des *Happy Hours* ont lieu hebdomadairement et consistent à cumuler le plus de points possible durant ce laps de temps, avec bonus à la clé pour le vainqueur. Le *Marathon* est un défi : il faut compléter une cellule entière (soit 1 000 *cubes*) en un temps donné (actuellement 26 heures). Le *Trivia* est une sorte de QCM, le challenge du *Cube Diabolique* consiste à colorier le mieux possible des régions particulièrement difficiles du fait que les branches neuronales de différentes cellules sont étroitement enchevêtrées. Ces événements ponctuels sont des éléments qui pimement le jeu en introduisant compétition et suspense. Depuis mai 2014 les joueurs peuvent se constituer en équipes, ce qui donne

Jeux biologiques

lieu à un classement inter-équipes. Enfin, tout récemment (3 juin 2014) a été lancé *Lore of Kor*, un concours de coloriage de neurones sur fond de scénario de science-fiction, au cours duquel s'affrontent deux équipes, avec primes et bonus divers à la clé. Les développeurs font *vivre le jeu* ; ils sont donc très attentifs à compenser la monotonie de la tâche par toutes sortes de dispositifs censés encourager interactivité et compétitivité, tout en suscitant, entretenant et renforçant le sentiment d'appartenir à une communauté.

L'autre méthode de fidélisation et d'encouragement passe par une exploitation poussée du concept très porteur de *science citoyenne*, qui se traduit par une valorisation extrême du joueur en tant qu'acteur de la recherche. Les chercheurs et développeurs communiquent fréquemment sur le blog du jeu et ne manquent pas une occasion d'associer les joueurs, sous l'appellation *The Eyewirers*, aux avancées et résultats obtenus [Kim *et al.*, 2014]. L'autre méthode de fidélisation et d'encouragement consiste à *caresser le joueur dans le sens du poil* et renforcer, *via* l'octroi de titres et diplômes aux consonances ésotériques, son sentiment d'appartenance à la confrérie très fermée des défricheurs de réseaux neuronaux.

Design/ambiance : l'interface de jeu est soignée, claire, agréable. Le fond noir met en valeur les modélisations de neurones en 3D qui, avec tous leurs prolongements ramifiés donnent lieu à des figures assez spectaculaires que l'on peut aisément faire tourner dans toutes les directions. Sur la page d'accueil, un tableau récapitule les différents modes de classement (quotidien, hebdomadaire, mensuel, individuel, par équipe). Dans l'interface de jeu proprement dite, la vue du *cube* en 3D est manipulable et orientable, et il est assez plaisant de voir progressivement apparaître en 3D le volume de la branche neuronale que l'on est en train de colorier sur la coupe transversale en 2D.

1.7. Citizen sort

*Citizen sort*¹¹ est un portail de *science citoyenne* conçu par un groupe de chercheurs et d'étudiants de l'école des Sciences de l'Information de l'Université de Syracuse, dans l'état de New York. Les jeux proposés ont été développés en collaboration avec des biologistes et des naturalistes. L'objectif de cette initiative est double :

- générer *via* les jeux des données utiles aux biologistes/naturalistes (notamment en entomologie, botanique et océanographie) ;
- aider les chercheurs en sciences de l'information et les spécialistes dans le domaine des interactions homme-machine à comprendre comment interagissent les concepts de jeu, de science et de citoyenneté, et à évaluer le rôle de la ludification en tant que moteur de la motivation. Le portail *Citizen sort* permet d'étudier comment différents types de jeux, s'adressant à différents types de joueurs, peuvent générer différentes qualités de données.

¹¹*Citizen sort* est accessible à l'adresse <http://www.citizensort.org/>.

Jeux biologiques

Le portail *Citizen sort* propose des jeux pour classer diverses espèces de plantes, d'insectes, d'animaux. Deux jeux sont actuellement accessibles, *Happy match*¹² et *Forgotten Island*¹³ ; un troisième, *Mark With Friends*¹⁴, est en développement. Le portail *Citizen Sort* annonce *via* le classement général plus de 3 000 joueurs, mais seulement 1 200 actifs (avec un score positif).

1.7.1. Happy match

Type de tâche concernée : microtâche

But : il s'agit d'aider à la classification d'êtres vivants en caractérisant des spécimens photographiés suivant différents critères. Concrètement, une série de 5 à 10 photos est proposée, et le joueur va devoir, pour chaque photo, se prononcer sur plusieurs critères morphologiques, en assimilant le spécimen à une forme de référence, dont un exemplaire caractéristique est présenté au-dessous. Par exemple, s'il s'agit de papillons, il s'agira d'assimiler chaque photo à un type général en fonction de la forme des ailes au repos, puis de la couleur dominante des ailes, puis de la forme du motif des ailes, etc.

Les données morphologiques ainsi récoltées par *crowdsourcing*, concourent à identifier formellement l'échantillon au niveau spécifique, voire sub-spécifique. Comme par ailleurs les photos sont datées et possèdent l'information de localisation (par géolocalisation), la corrélation des informations taxonomiques avec la date et la localisation peut par exemple renseigner sur l'impact de certaines formes d'urbanisation sur un écosystème, ou mettre en évidence un changement climatique, une modification de la biodiversité, l'apparition d'une espèce invasive, etc. Actuellement il est possible de jouer avec des photos de papillons (*Happy Moths*), de requins (*Happy Sharks*), ou de raies (*Happy Rays*).

Intérêt du crowdsourcing : le *crowdsourcing* est utilisé ici pour faire accomplir une tâche très lourde en termes de temps et de volume de données à traiter. Comme souvent dans le cas de micro-tâches, c'est la redondance et l'agrégation des données fournies par les joueurs qui, à condition qu'ils soient suffisamment nombreux, vont fournir des données d'aussi bonne qualité que les annotations d'experts.

Compétences humaines requises : elles sont minimales et se résument à un sens de l'observation, un intérêt pour la nature... et de la patience...

Education du joueur : en jouant de manière répétée, et en utilisant les éléments de correction fournis (voir challenge/ludification), on peut exercer son œil et apprendre à bien discriminer certains

¹²*Happy match* est accessible à l'adresse <http://www.citizensort.org/web.php/happymatch>.

¹³*Forgotten Island* est accessible à l'adresse <http://www.citizensort.org/web.php/forgottenisland>.

¹⁴*Mark With Friends* est accessible à l'adresse <http://www.citizensort.org/web.php/markwithfriends>, mais pas encore effectif.

Jeux biologiques

des éléments morphologiques qui sont à la base des clés de détermination, et devenir incollable sur les noms latins de famille, de genre et d'espèce. On peut raisonnablement supposer que ce genre de jeu plaît à des personnes qui ont déjà des compétences en taxonomie, comme des entomologistes amateurs par exemple.



Figure 1.9.* Citizen Sort : Happy match. *Le premier écran du jeu présente 10 photos de papillons. A ce stade le joueur doit juste valider les photos, c'est-à-dire éliminer celles qui seraient floues, tronquées, ou trop mauvaises être exploitées. Dans les écrans suivants, il classera les papillons suivant divers critères morphologiques (forme des ailes, couleur dominante, etc.).*

Interactivité : elle n'est pas du tout développée dans *Happy Match*. Le joueur caractérise ses images dans son coin, sans contact avec ce que font les autres joueurs, sans espace de discussion type messagerie instantanée ou forum. Par contre, dans le portail *Citizen Sort* figure l'annonce d'un jeu à venir, *Mark With Friends* qui semble, comme son nom l'indique, mettre l'interactivité au centre de la stratégie d'acquisition de données.

Challenge/ludification/interface : l'interface est visuellement agréable et on sent qu'elle a été conçue avec le concours de designers. Par rapport à d'autres jeux d'annotation d'images (voir ceux du portail *Zooniverse*), on note un effort très net pour développer les aspects ludique et esthétique, dès l'interface, qui est organisée pour présenter la tâche à réaliser de façon harmonieuse et ergonomique. Habilement, l'élément essentiel de ludification est également le paramètre qui permet d'apprécier la qualité des observations et donc de valider les données fournies par le joueur : dans la série d'images proposées sont *cachées* une ou deux photos d'individus déjà connus et identifiés avec

Jeux biologiques

certitude. Si le joueur classe correctement ces deux images sur tous les paramètres morphologiques proposés, on peut raisonnablement penser que le classement qu'il opère sur les autres images est fiable. A la fin de la partie, le résultat révèle si les deux images connues ont été bien classées sur tous les paramètres ; dans ce cas le joueur gagne des points *bonus* et les deux espèces apparaissent en tant que spécimens collectés dans son profil ; dans le cas contraire, pour chaque caractéristique morphologique mal attribuée, apparaît la décision correcte, en regard de la photo du spécimen, ce qui lui permet de savoir où et pourquoi il s'est trompé, et ajoute donc un élément pédagogique à l'aspect ludique.

1.7.2. Forgotten Island

Type de tâche concernée : microtâche

Forgotten Island est la version jeu d'aventure de *Happy Match*. La tâche demandée est la même (classer des images d'animaux suivant des critères morphologiques), mais dans un contexte qui imite les jeux d'aventures : on joue le rôle d'un aventurier égaré sur une île où une explosion a ravagé un laboratoire et dispersé des spécimens d'animaux, qu'il faut répertorier et reclasser. Le personnage que l'on incarne se déplace dans un décor et peut glaner des objets nécessaires à sa mission, mais aussi se téléporter d'un endroit stratégique de l'île à un autre, sous la férule d'un robot qui lui dit quoi faire et où aller...

1.8. Le projet Nightjar

But/généralités : le *Projet Nightjar*¹⁵ est un programme de recherche lancé en août 2013, par le Groupe d'écologie sensorielle de l'Université d'Exeter, en collaboration avec le Groupe d'écologie comportementale de l'Université de Cambridge. La thématique concerne l'étude des stratégies de camouflage dans la nature, en relation avec la survie, et le modèle est celui de l'engoulevent (*nightjar*, en anglais), un oiseau, qui niche et pond au sol. Des photographies d'engoulevents dans leur milieu naturel sont modifiées en fonction des paramètres visuels de leurs prédateurs naturels (différents des nôtres) pour comparer l'aspect des œufs, des poussins, et des adultes à celui de leur environnement naturel afin de quantifier l'efficacité du camouflage. L'étude porte sur diverses espèces d'engoulevents, toutes très dépendantes de stratégies de camouflage appropriées pour échapper à leurs prédateurs. D'autres oiseaux sont également étudiés (pluviers et coursiers), mais les engoulevents constituent un modèle de prédilection car les œufs, les adultes, et les poussins témoignent de stratégies de camouflage différentes. L'équipe de chercheurs, en collaboration avec un concepteur

¹⁵Le *Projet Nightjar* est accessible à l'adresse <http://nightjar.exeter.ac.uk/>.

Jeux biologiques

de jeu, a développé trois jeux pour tester l'efficacité des stratégies de camouflage des œufs et des adultes d'engoulevents, en fonction du type de perception visuelle de leurs divers prédateurs.



Figure 1.10.* Le projet Nightjar. *L'écran d'introduction du Nightjar game, un des jeux du projet*

Intérêt du crowdsourcing pour le problème à résoudre : le problème à résoudre est l'évaluation de l'efficacité de la stratégie de camouflage, qui peut être mesurée par le délai moyen de repérage d'un nid ou d'un oiseau dans un décor. Le joueur va choisir d'incarner une catégorie de prédateur et chercher à repérer le nid ou l'oiseau en visionnant des photographies prises dans les sites naturels. Ces photographies ont été modifiées de manière à correspondre à la vision di-, tri-, ou tétrachromatique caractéristique de chacun des trois types de prédateurs naturels de l'engoulevent. Les données récupérées par *crowdsourcing* sont donc des délais de repérage, qui sont étroitement corrélés avec l'efficacité du camouflage.

Interactivité/éducation du joueur : le type de jeu proposé ne se prête pas à interagir avec d'autres joueurs. Cependant on note un effort louable des concepteurs pour expliquer la nature de leurs travaux : des informations scientifiques précises et détaillées sur le projet sont facilement acces-

Jeux biologiques

sibles *via* le site du projet, tandis que des données plus techniques concernant les jeux sont fournies *via* le blog du concepteur de jeu¹⁶.

1.8.1. Nightjar game/Nest game

Type de tâche concernée : ni microtâche, ni macrotâche

But : *Where is that nightjar ? (Nightjar game)*¹⁷ est le premier jeu lancé en août 2013, suivi en novembre par *Where is that nest ? (Nest game)*¹⁸, qui fonctionne sur le même principe. Le joueur doit, en un temps limité, repérer soit un oiseau, soit un nid sur la photo qui lui est présentée, après avoir choisi de quel type de prédateur (singe = vision trichromatique, ou mangouste = vision dichromatique), il avait la vision. Le délai avec lequel le nid ou l'oiseau est repéré (ou pas) est directement corrélé avec l'efficacité de la stratégie de camouflage.

Challenge/ludification : une partie consiste à chercher un oiseau ou des œufs dans 20 photos successives, après avoir accepté que les données de la partie soient exploitées dans le cadre d'un travail scientifique, indiqué sa tranche d'âge, et choisi quel prédateur on incarnait (ce qui conditionne les paramètres colorimétriques des photos présentées). Le principal élément ludique est la tâche elle-même : il est amusant de rechercher un oiseau (ou des œufs) camouflé(s) dans un décor naturel, et pas forcément facile de le repérer quand on voit le paysage dans un camaïeu de vert, c'est-à-dire avec la vision dichromatique d'une mangouste. Le temps limité (30 s) pimente la tâche. Si on clique à un endroit où il n'y a rien, une grosse croix rouge apparaît accompagnée d'un son grave. Si on n'a rien trouvé au bout du temps imparti, la solution apparaît sous la forme d'un rectangle autour de l'élément camouflé, et l'échec est salué d'une mention humoristique « Vous aurez faim, ce soir ! ».

1.8.2. Egglab game

Type de tâche concernée : ni microtâche, ni macrotâche

Audience/popularité : le blog du concepteur de jeu¹⁶ fait état, en mai 2014, de 9000 joueurs, qui ont fait 20 000 parties et testé 400 000 motifs de coquille représentant plus de 30 générations d'œufs artificiels. Le délai global moyen de repérage d'un œuf a augmenté d'environ 0,4 seconde depuis le début de l'expérience, initiée en avril 2014, la date de lancement ayant été opportunément choisie en relation avec Pâques !

Principe et but : *Egglab* est un logiciel conçu pour simuler la variabilité génétique qui caractérise les motifs visibles sur les coquilles d'œufs, lesquels jouent un rôle crucial dans le camouflage et

¹⁶<http://www.pawfal.org/dave/blog/>.

¹⁷*Nightjar game* est accessible à l'adresse http://nightjar.exeter.ac.uk/story/nightjar_game.

¹⁸*Nest game* est accessible à l'adresse http://nightjar.exeter.ac.uk/story/nest_game.



Figure 1.11.* Nightjar game. Un engoulevent (désigné par la flèche) est caché dans ce décor, et le joueur a 30 secondes pour le repérer et cliquer sur lui

sont donc directement corrélés à la prédation. Les œufs simulés sont des images qui sont juxtaposées aux photos des différents sites de ponte, et testés *via* le *Egglab game*¹⁹ : le joueur a 30 secondes pour localiser l'œuf dans le décor, et les motifs les moins repérables sont réinjectés dans le programme pour générer une nouvelle série de décors aboutissant à des coquilles de plus en plus mimétiques du décor ambiant, à l'image de ce qui se passe dans la nature *via* le processus de sélection naturelle.

Challenge/ludification : le jeu est bâti sur le même principe que les deux précédents, à la différence près que ce sont des œufs de synthèse (donc des dessins) qu'il faut repérer dans un décor naturel, qui est une photo de l'habitat d'une des trois espèces d'engoulevents étudiés. La présentation et le décor sont plus *cartoon*, on est invité à rentrer son nom en cliquant successivement sur des œufs portant des lettres, à la fin de la partie le temps moyen clignote, et le joueur est prévenu que grâce à lui, les œufs vont devenir de plus en plus difficiles à déceler. Il est invité à continuer à jouer pour voir comment la difficulté va évoluer, ce qui est une manière de tenter de le fidéliser. Une manière habile d'attirer un plus grand nombre de joueurs est la possibilité récente, pour le joueur, de cacher

¹⁹*Egglab game* est accessible à l'adresse <http://nightjar.exeter.ac.uk/egglab/>.

Jeux biologiques

lui-même des œufs dans un décor (*Making an Easter egg hunt*) et d'envoyer sa chasse aux œufs de Pâques (*Easter egg hunt*) à ses amis. Pour ce faire, il est invité à choisir un décor naturel (photo), puis à y glisser-déposer des œufs virtuels avec divers décors de coquilles, en choisissant lui-même l'endroit du décor où le camouflage lui semble optimal. Quant c'est terminé, un clic sur *send* génère une URL qu'il est possible de copier et envoyer par mail.

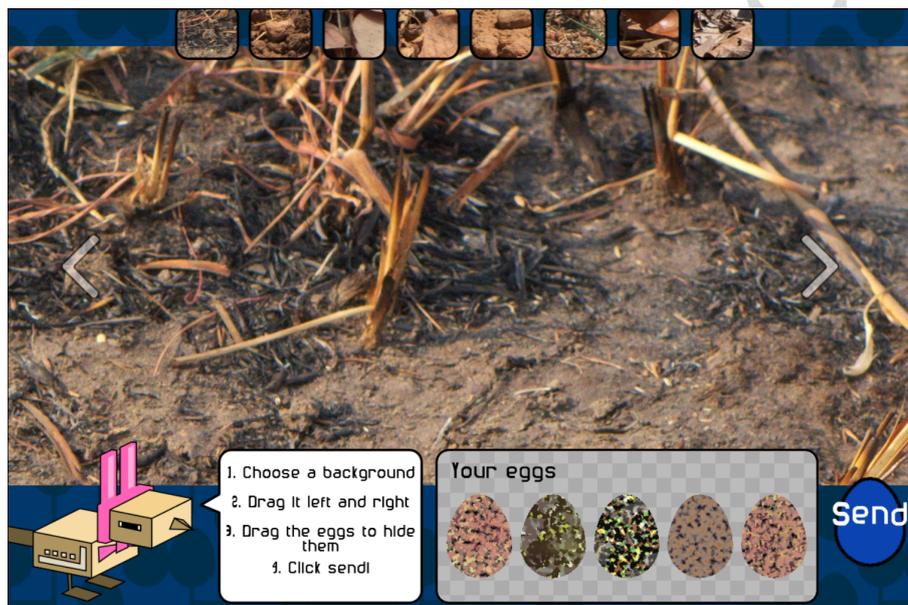


Figure 1.12.* Le jeu Egglab : making an Easter egg hunt (*Organisez une chasse aux œufs*) est une variante du jeu Egglab : le joueur choisit un décor, puis cache lui-même les œufs en les cliquant-déposant aux endroits de son choix. Il peut ensuite envoyer l'url du jeu créé à ses amis

Notes

¹<http://www.pri.org/stories/2014-05-12/your-video-games-may-revolutionize-health-and-scientific-research>.

²<http://boinc.bakerlab.org>.

³<http://singularityhub.com/2011/09/22/in-3-weeks-video-gamers-defeat-biochemical-puzzle-that-scientists-couldnt-solve-for-years/>.

⁴<http://www.pri.org/stories/2014-05-12/your-video-games-may-revolutionize-health-and-scientific-research>.

⁵<http://biomedicalcomputationreview.org/content/biology-game-crowd>.

⁶<http://www.pourquoidocteur.fr/Jouez-pour-aider-la-recherche-sur-les-maladies-genetiques---4090.html>.

CHAPITRE 2

Jeux à visée médicale

Le secteur médical n'est pas en reste quand il s'agit de recourir au *crowdsourcing* via le jeu, que ce soit au niveau de la recherche et de la mise au point de techniques innovantes de traitement, ou de l'aide au diagnostic. Nous avons écarté délibérément une des activités de *crowdsourcing* à finalité médicale parmi les plus populaires, *Cell Slider* [Good et Su, 2013], proposée par le portail *Zooniverse*, car elle est dépourvue de tout aspect ludique, même vaguement esquissé. On notera que parmi les jeux décrits ci-dessous, qu'il s'agisse de microtâche ou de macrotâche, les compétences requises de la part des joueurs sont très inégales et vont d'une simple aptitude à l'observation (*Worm Watch Lab*, *Malaria Training Game*), à une culture scientifique, génétique, ou médicale poussée voire très poussée (*Dizeez*, *The Cure*) en passant par la dextérité et la rapidité typiques des adeptes de jeux sur console (*Play to Cure*).

2.1. Nanodoc

Type de tâche concernée : plutôt macrotâche, mais des difficultés subsistent dans l'utilisation.

But/lancement/auteurs : *Nanodoc*¹ est un jeu lancé en septembre 2013 par Sangeeta Bhatia, chercheur au Koch Institute for Integrative Cancer Research du MIT, et son équipe. Il concerne la nanomédecine et a pour but la conception et la mise au point de traitements à base de nanoparticules. Les nanoparticules sont actuellement étudiées dans de nombreux laboratoires de recherche à travers

¹*Nanodoc* est accessible à l'adresse <http://nanodoc.org/>.

Jeux médicaux



Figure 2.1.* Nanodoc. Il faut fabriquer à l'aide de l'éditeur de nanoparticules, la combinaison qui permettra de remplir la mission indiquée dans la figure gauche. A droite, le résultat (succès) : toutes les cellules cancéreuses (en gris) ont été détruites

le monde car elles représentent un espoir considérable pour traiter les cancers, et, d'une manière générale, toutes les maladies qui résistent aux thérapies classiques. Un des intérêts majeurs des nanoparticules est leur capacité à délivrer le traitement sur le site même de la tumeur : on peut leur donner une taille idéale pour qu'après injection, elles circulent dans les vaisseaux sanguins et ne puissent s'en échapper qu'au niveau des vaisseaux plus lâches qui irriguent les tumeurs cancéreuses. De ce fait, les nanoparticules porteuses de médicaments vont s'accumuler passivement au niveau des tumeurs tout en évitant d'endommager les tissus sains.

Audience : une source fait état de 50 000 simulations effectuées par 2600 joueurs en octobre 2013, soit 1 mois après le lancement⁷. Il est impossible, dans l'interface du jeu, de trouver le nombre de joueurs inscrits.

Intérêt du crowdsourcing dans le problème à résoudre : il y a de multiples manières de concevoir des nanoparticules et de multiples paramètres qui influencent leur comportement. Le défi est de comprendre ce qui, dans le design de la particule, va permettre, en modifiant son mode d'action et ses interactions avec les autres particules et l'environnement physique de la zone tumorale, d'optimiser la délivrance et l'efficacité du traitement. L'objectif est également de faire en sorte que les nanoparticules agissent à la façon d'un système organisé (tel qu'il en existe dans la nature chez les insectes sociaux, par exemple) afin d'améliorer de façon synergique leur effet thérapeutique.

Il y a donc des milliers de combinaisons, de modes d'action, de stratégies thérapeutiques possibles, et il faudrait un temps considérable pour les examiner et réaliser les tests et essais nécessaires pour pouvoir les valider. De plus, chaque problème est différent, ce qui exclut la programmation d'un ordinateur qui concevrait automatiquement des nanoparticules. C'est pourquoi l'équipe de Sangeeta Bhatia a conçu un simulateur qui permet de modéliser comment les nanoparticules interagissent

Jeux médicaux

entre elles et avec la tumeur. L'idée est de faire de ce simulateur une interface de jeu : d'un côté les bioingénieurs de l'équipe de recherche conçoivent divers scénarios de tumeurs, tandis que de l'autre les joueurs vont imaginer diverses stratégies de traitement en concevant des nanoparticules virtuelles qui seront testées *via* le simulateur.

L'idée est que les stratégies les plus prometteuses soient validées par l'équipe de recherche pour servir de base à des tests *in vitro*.

Compétences humaines requises : esprit d'initiative, curiosité, inventivité, ténacité, patience. . .

Education du joueur : dès le début du jeu, le joueur est instruit *via* un tutoriel sur les différents paramètres inhérents à la création et à l'utilisation des nanoparticules (concentration utile, spécificité, taille des particules, ciblage. . .) et amené à tester divers assemblages et configurations de manière à tuer toutes les cellules tumorales en laissant intactes les cellules saines. Chaque paramètre est expliqué sommairement, puis le joueur est invité à le mettre en pratique en fabriquant un traitement capable de remplir l'objectif proposé. Le problème vient de ce que les explications sont trop sommaires et le joueur se retrouve à faire des dizaines de tentatives sans forcément trouver la bonne combinaison. Étant donné qu'à ce stade, le joueur est encore dans le tutoriel, il pourrait s'attendre à ce qu'en cas d'échecs répétés, au lieu de la formule « Vous n'avez tué aucune cellule cancéreuse. Essayez d'améliorer le traitement », la solution lui soit donnée. En cas d'échec de traitement, il est tout de même possible de passer au niveau suivant, mais c'est extrêmement frustrant et le problème est récurrent. La rubrique *learn* du jeu, même si elle apporte beaucoup d'informations assez bien vulgarisées sur les notions biochimiques réelles impliquées dans les différentes stratégies de traitement illustrées par le tutoriel, n'aide absolument pas le joueur à comprendre pourquoi ce qu'il tente ne fonctionne pas. Décourageant. . .

Interactivité : de toute évidence, *Nanodoc* est un jeu solitaire, la compétition avec les autres joueurs n'est pas prévue, les possibilités d'interaction se limitent à un blog où, à part la présentation du jeu, on ne trouve aucun commentaire. La rubrique *learn* se présente également comme un blog, et les commentaires se comptent sur les doigts d'une main.

Challenge/émulation/ludification/design : la ludification consiste essentiellement à appeler le joueur *Docteur*, à le féliciter quand il a réussi et à lui attribuer un *certificat* mentionnant les compétences acquises au fur et à mesure du tutoriel. Pas d'émulation puisqu'on joue seul, pas d'autre satisfaction que celle d'avoir atteint l'objectif et détruit les cellules tumorales en préservant les saines. Le design du jeu est par contre plutôt réussi, l'interface sobre, l'animation illustrant l'impact du traitement plaisante bien qu'un peu longue, mais c'est sans doute délibéré et destiné à créer un peu de suspense : « va-t-on réussir à éliminer toutes les cellules tumorales ? »

Conclusion : le jeu est prometteur mais décevant, le tutoriel, mal fait, manque complètement son but puisqu'on est bloqué en pleine phase d'apprentissage et découragé avant de rentrer dans le

Jeux médicaux

vif du sujet. Pas de classement, pas d'interaction avec les autres joueurs, pas d'émulation... On est loin de l'ambiance de *Eyewire*, alors que paradoxalement la tâche à réaliser semble infiniment plus variée et intéressante.

2.2. Dizeez

Type de tâche concernée : microtâche, mais réservé aux spécialistes.

Généralités/présentation : *Dizeez*², mis en ligne en octobre 2011 par The Su Lab, un laboratoire californien spécialisé dans l'adaptation des outils informatiques aux problématiques de la recherche biologique, lance un défi à ceux qui pensent avoir des connaissances sur les bases génétiques des maladies humaines. Le principe est simple : le nom d'un gène est donné (par exemple : IYD (iodotyrosine deiodinase)), et il faut l'associer à l'une des 5 maladies proposées, en temps limité. Chaque réponse juste donne des points.

Problème à résoudre et intérêt du crowdsourcing : en général, dans les bases de données, les liens entre gènes et maladies sont corrects. On s'attend donc, en analysant les données issues du jeu, à ce que les réponses des joueurs confirment ce qui est déjà établi. Mais si suffisamment de personnes jouent, on peut également supposer que seront ainsi révélés des liens gène-maladie absents des bases de données. Le jeu a ainsi pour vocation, en comblant des lacunes dans les bases de données d'annotation de gènes, de réduire le fossé entre la connaissance disponible dans la littérature biomédicale et celle des banques de données, généralement incomplètes. D'après les concepteurs, un mois après son lancement, *Dizeez* a déjà abouti à l'identification de nouveaux liens gène-maladie qui sont bien établis dans la littérature, mais ne peuvent être pris en compte dans les analyses informatiques bio-statistiques puisqu'ils ne figurent pas encore dans les bases de données.

Les concepteurs se disent conscients de la nécessité d'apporter des améliorations au jeu, mais très excités par les potentialités de ce type d'approche (le GWAP) pour aider à structurer la connaissance. Cependant ils n'ont publié aucun nouveau billet dans le blog associé, depuis mars 2012.

Compétences humaines requises : ce jeu nécessite une connaissance fine du génome et des associations gènes/maladies.

Education du joueur : le joueur n'apprend rien, il doit au contraire... savoir déjà pour jouer efficacement ! Ne pas connaître les gènes par leur dénomination scientifique est un handicap et réduit l'expérience à un simple jeu de hasard.

Ludification/interface/critique : beaucoup des ingrédients susceptibles de rendre un jeu addictif sont présents (temps limité, système de gain/perte de points, classement, top 10), l'interface est d'une sobriété efficace, sans fioritures, peu de couleurs, on sent qu'on n'est pas là pour rigoler.

²*Dizeez* est accessible à l'adresse <http://sulab.scripps.edu/dizeez/>.

Jeux médicaux

Mais... ce jeu s'adresse à... des médecins, des généticiens, des acteurs de la recherche sur les maladies génétiques, des biochimistes, certainement pas au grand public... Qui, dans le grand public, même s'il a des connaissances en biologie, voire en génétique, peut savoir que le gène HOXB7 est associé à la leucémie, ou le gène BCL10 au lymphome ? *Dizeez* est un jeu fait par des spécialistes pour des spécialistes... Une version élitiste de *crowdsourcing* ne concernant effectivement qu'un nombre restreint de joueurs.

2.3. The Cure

Type de tâche concernée : inclassable, plutôt à tendance macrotâche, destiné aux initiés.

Principe/description : l'objectif du jeu est de bâtir un modèle efficace de prédiction de durée de survie au cancer du sein basé sur l'expression génique et l'information issue d'échantillons tumoraux. *The Cure*³, développé et lancé en septembre 2012 par le même laboratoire que *Dizeez* est présenté comme un jeu de poker dans lequel les cartes sont représentées par des gènes : il s'agit d'assembler la meilleure *main* en choisissant les gènes qui la composent pour battre Barney, un automate qui propose une combinaison calculée d'après les données relatives aux taux de survie connus. Chaque *carte* (gène) ajouté fait varier le score et à terme, l'assemblage de 5 gènes donne un score plus ou moins élevé que celui de Barney. Le but est d'utiliser les données génétiques relatives à des cas cliniques réels et de les corrélérer à ce que l'on sait de l'évolution de la maladie chez ces patients pour en tirer des modèles prédictifs et des informations sur l'implication de tel gène ou de telle combinaison de gènes dans la capacité d'une tumeur à métastaser, par exemple.

Audience : entre septembre 2012 et septembre 2013, 1 000 joueurs se sont inscrits et ont joué presque 10 000 parties, d'après [Good *et al.*, 2014].

Intérêt du crowdsourcing pour le problème à résoudre : chercher des modèles prédictifs se heurte à deux problèmes : la reproductibilité incertaine, due à la variabilité biologique naturelle, mais surtout l'impossibilité de tester toutes les combinaisons de gènes possibles en raison de leur trop grand nombre : une combinaison unique de 25 gènes peut constituer un excellent modèle prédictif, mais avec plus de 20 000 gènes on peut avoir de l'ordre de 10^{82} combinaisons possibles de 25 gènes. D'où l'intérêt de recourir à l'intelligence et à la connaissance humaines pour élaborer des combinaisons associant les gènes non pas au hasard comme le ferait une machine, mais en se guidant par les connaissances établies sur les gènes, les mécanismes biologiques impliqués dans les cancers, et les données cliniques sur les maladies. Le joueur est ainsi invité à composer ses *mains* à partir de ses connaissances s'il en a, ou en recherchant le rôle de tel ou tel gène sur le Web. Pour faciliter quelque peu le choix, une information sur chaque gène, quand toutefois elle est connue, est

³*The Cure* est accessible à l'adresse <http://genegames.org/cure>.

Jeux médicaux

disponible dans l'interface de jeu. Les combinaisons élaborées par les joueurs vont alimenter des algorithmes d'apprentissage qui permettront de développer des modèles prédictifs sur le pronostic du cancer du sein.

The screenshot shows the 'The Cure' game interface. At the top, there is a pink navigation bar with 'The Cure', 'Help!', 'Contact', 'Advanced Cure', and 'Logout'. Below this, a hand of genes is displayed in a central area, with genes arranged in a pattern similar to a poker hand. The genes shown are: NEK2, SFPQ, ST6GALNA, CEBPA, RET, MYLK, PIK3R1, GINS1, ERBB2, TFF1, HNRNPL, FAM222B, CSF1R, SAP30, FABP4, KIT, PDGFRA, LIMCH1, CTSK, FGFR2, DPY19L4, OPN3, FGFR3, PSD3, and CDC37. To the right of the hand, there is a window titled 'Barney's score: 56' and 'Your score: 61'. Below the scores, there is a search bar and a window displaying information for the gene 'CTSK' (cathepsin K). The window includes a description of the protein, its function, and a reference to RefSeq.

Figure 2.2.* The Cure. Par analogie avec le poker, il faut constituer une main en choisissant 5 gènes dans la partie centrale. Les gènes choisis s'affichent dans la fenêtre du haut, et le texte de droite donne des informations sur le gène survolé par la souris, afin de guider le choix. Chaque gène choisi fait progresser les scores, le but est de battre celui de Barney (l'algorithme de choix).

Compétences humaines requises : jouer nécessite une certaine culture en génétique/médecine/biochimie, qui semble être difficilement compensable par la curiosité nécessaire pour aller chercher sur le Web les informations relatives aux gènes connus pour leur implication dans certains cancers, afin de constituer ses *main*s en connaissance de cause. Dans la FAQ, il est tout de même précisé que plus on aura de connaissances en génétique et sur le cancer, mieux ce sera pour jouer, mais qu'on peut aussi les acquérir au fil du jeu, puisque beaucoup d'informations scientifiques sont fournies dans le jeu, ainsi que des liens vers des ressources Web.

Jeux médicaux

Education du joueur : le joueur apprend en jouant... s'il a la patience de lire et d'assimiler toutes les informations relatives aux gènes qui lui sont proposés dans l'interface de choix. Mais c'est un peu contradictoire avec le principe du jeu de poker, qui suppose d'enchaîner les parties rapidement et de faire (au moins partiellement) confiance au hasard. Et le choix de gènes *au hasard* pour constituer la *main* donne plutôt de bons résultats puisque en procédant de la sorte, on bat Barney très souvent. Cependant, les concepteurs affirment qu'en testant les modèles prédictifs issus des données de jeu sur des jeux de données réelles, il est facile de distinguer les joueurs sérieux de ceux qui ont assemblé des gènes au hasard. De plus, les modèles prédictifs issus du jeu démontreraient clairement l'acquisition de compétences chez les joueurs, et donnent des résultats comparables à ceux obtenus par les méthodes classiques [Good *et al.*, 2014].

Ludification/Design : l'interface de jeu est sobre et agréable, l'ergonomie plaisante. Chaque partie est rapide, et l'évolution des deux scores au fur et à mesure de l'adjonction de gènes entretient le suspense. Les scores cumulés donnent lieu à un classement inter-joueurs.

2.4. Malaria Training Game

Type de tâche concernée : microtâche

But : lancé le 25 avril 2014, journée mondiale du paludisme, par Biogames, département du laboratoire de recherche Ozcan de l'UCLA spécialisé dans la production de jeu pour le télédiagnostic médical, *Malaria Training Game*⁴ vise à faire reconnaître par *crowdsourcing* les cellules infestées par la malaria. Il s'agit en fait d'un module de formation au cours duquel le joueur est invité à examiner des photos de cellules individuelles isolées et caractériser chacune d'elle comme *saine*, *infestée* ou *douteuse*. Chaque partie consiste à diagnostiquer environ 500 cellules, qui ont déjà été caractérisées par des professionnels. Un score est attribué à la fin en fonction du nombre de cellules correctement désignées, mais aussi du nombre de *faux positifs*, de *faux négatifs* et de cellules caractérisées à tort comme *douteuses*. La dimension éducative du jeu est de montrer au joueur pour quelles cellules et dans quel sens (*faux négatif*, *faux positif*, *faux douteux*) il s'est trompé, afin qu'il puisse s'améliorer à la partie suivante.

Le but à long terme est double : 1) que ce module ludique de formation soit largement utilisé pour améliorer la formation du personnel médical et la qualité du diagnostic dans toutes les zones où sévit le paludisme et 2) qu'il serve à instruire des algorithmes d'apprentissage dans le but d'automatiser le diagnostic.

L'intérêt de ce type d'approche est qu'elle est adaptable à d'autres maladies, et qu'elle peut également servir à mesurer la parasitémie chez des patients sous traitement.

⁴*Malaria Training Game* est accessible à l'adresse <http://biogames.ee.ucla.edu/training>.

Jeux médicaux

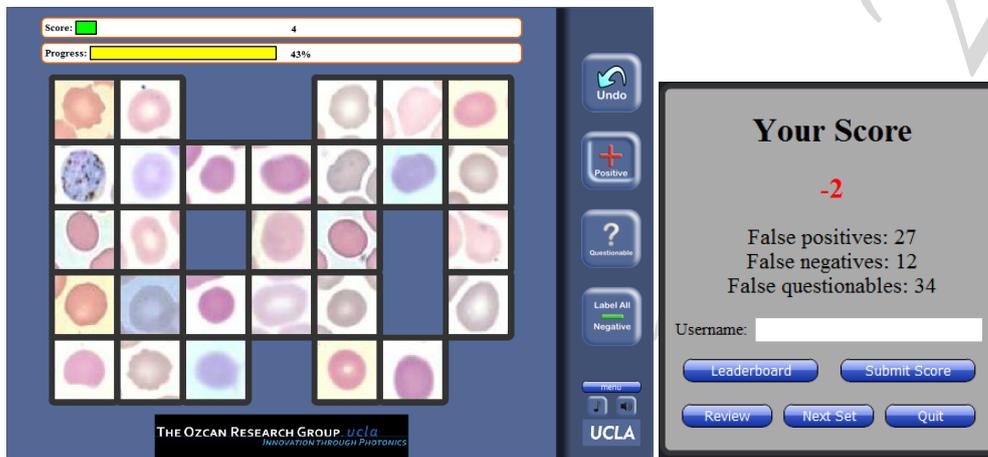


Figure 2.3.* Malaria Training Game. *Figure gauche* : on caractérise les cellules (positive, négative, douteuse) en cliquant dessus après avoir choisi l'outil approprié dans le menu de droite. La figure droite montre le score à l'issue de la partie

Intérêt du crowdsourcing pour le diagnostic de la malaria : un pathologiste, pour établir un diagnostic fiable de la malaria, doit inspecter au moins 1000 globules rouges. C'est une tâche longue et fastidieuse, où l'erreur est fréquente : d'après les concepteurs, 60 % des cas déclarés sont des *faux positifs*. Toute réduction du nombre de *faux positifs* peut éviter traitements et hospitalisations inutilement coûteux. Dans leurs travaux publiés, les concepteurs estiment avoir montré qu'en exploitant les qualités visuelles et les capacités d'apprentissage du public *via* des jeux qui combinent phase d'apprentissage et feed-back, on pouvait parvenir à un diagnostic fiable sur des échantillons médicaux numériques. En cas de diagnostic binaire (infesté/sain) on peut approcher la précision et la fiabilité des experts, puisque les diagnostics établis *via* le jeu ne diffèrent que de 1,25 % de ceux des professionnels qualifiés [Mavandadi *et al.*, 2012].

Compétences humaines requises/éducation : sens de l'observation, volonté de s'améliorer en utilisant la correction pour dépister et éliminer les sources d'erreurs. Il est évident qu'en pratiquant le jeu, et en partant d'une ignorance absolue, on s'améliore rapidement. Le jeu ne nécessite aucune connaissance préalable, en particulier aucune compétence autre que celle de savoir reconnaître une hématie infectée.

Challenge /ludification/ergonomie : le jeu ne nécessite pas d'inscription. Une partie (500 cellules) prend quelques minutes, c'est suffisamment rapide pour qu'on ait envie de tenter d'améliorer son score en recommençant. Les cellules sont présentées sur une grille de 5 sur 7, soit 35 cellules à caractériser par écran. La caractérisation des cellules est intuitive, elle se fait par simple clic après

Jeux médicaux

avoir sélectionné l'outil adapté. Deux barres de progression indiquent, l'une le score, en vert si les bonnes réponses prédominent, en rouge sinon, l'autre le pourcentage de cellules examinées. L'interface est simple, claire, sans fioritures, deux boutons permettent de régler l'ambiance sonore du jeu. A l'issue de la partie, le joueur peut prendre connaissance de ses erreurs, et soumettre son score afin d'apparaître dans le classement.

2.5. Malaria Spot Game

Type de tâche concernée : microtâche, avec pour slogan : « un jeu d'une minute contre la malaria ».

Audience : un mois après le lancement du jeu (avril 2012), des joueurs de 95 pays ont joué plus de 12 000 parties et généré une base de données de plus de 270 000 clics sur les images de test. Selon [Luengo-Oroz *et al.*, 2012], en avril 2013, plus de 15 000 personnes de 100 pays différents ont généré presque 1 000 000 de clics.

But : développé par Miguel Luengo-Oroz et son équipe (Biomedical Image Technologies Group & CEI Moncla UPM-UCM) de l'Université polytechnique de Madrid, *Malaria Spot Game*⁵ exploite une approche un peu différente de celle de *Malaria Training Game* pour utiliser le *crowdsourcing* à des fins de diagnostic. Le jeu, décrit dans [Luengo-Oroz *et al.*, 2012] semble être toujours en phase de test et constitue l'étude préliminaire à la mise en place d'une plate-forme mondiale spécialisée dans le diagnostic en ligne, de la malaria dans un premier temps, mais potentiellement de n'importe quelle maladie dont le dépistage repose sur l'analyse d'images.

Actuellement, le diagnostic de la malaria repose sur la détection et le comptage des parasites sur une lame de frottis sanguin. Il s'agit d'une tâche manuelle, qui prend environ 30 minutes à un spécialiste. D'où l'intérêt de mettre au point un système de dépistage plus rapide, aussi performant, facilement accessible. La couverture mondiale de la téléphonie mobile et le web représentent des opportunités de pouvoir rendre des images accessibles partout dans le monde. Or, après en avoir vu quelques-uns, il est assez facile d'identifier les parasites sur une image de frottis sanguin. Le principe du jeu est donc de soumettre aux joueurs des images de frottis sanguins et de leur demander de parcourir l'image en cliquant sur toutes les formes qu'ils assimilent à un *plasmodium*. Les joueurs réalisent ainsi un comptage, le but étant de trouver tous les parasites présents sur l'échantillon.

Intérêt du crowdsourcing pour le problème à résoudre : de l'analyse de tous les clics des joueurs, qu'ils soient justes ou faux on va pouvoir 1) évaluer la rapidité d'un non-spécialiste et la précision avec laquelle il est capable de détecter des agents pathogènes sur une image ; 2) mettre au

⁵*Malaria Spot Game* est accessible à l'adresse <http://www.malariaspot.org/game/>.

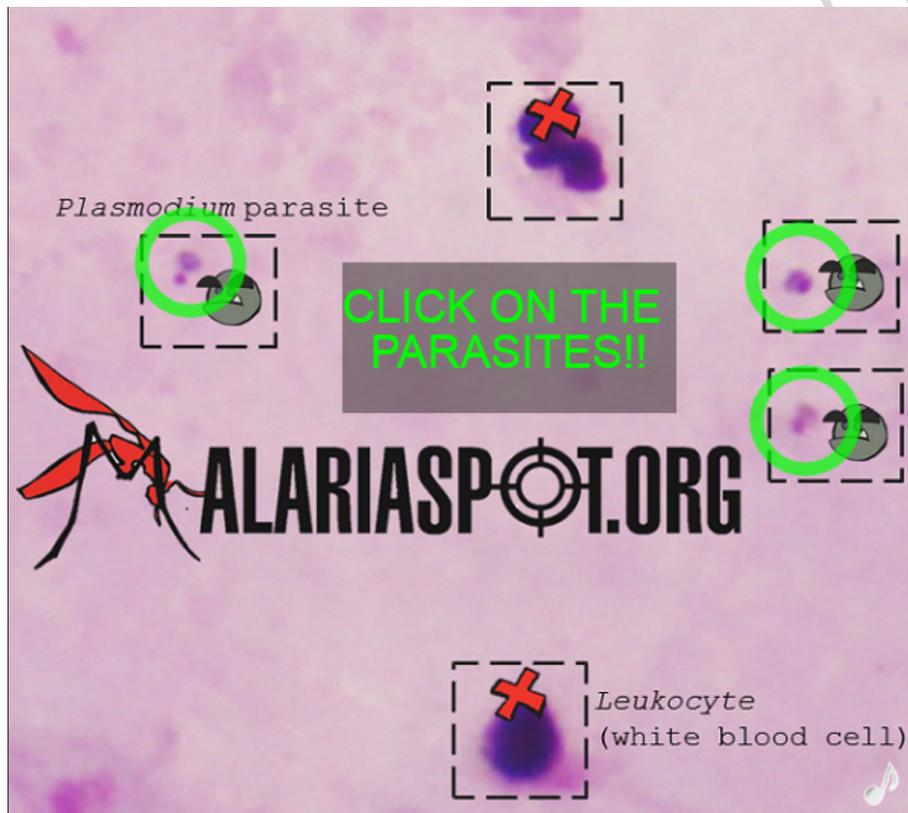


Figure 2.4.* Malaria Spot Game. *L'écran d'accueil du jeu fait office de tutoriel : des cercles verts désignent les formes correspondant à des parasites, qu'il faudra découvrir, tandis que les croix rouges indiquent des structures cellulaires qu'il ne faut pas confondre avec un plasmodium*

point une manière de combiner les résultats des différents joueurs pour obtenir un diagnostic aussi précis que celui d'un expert qualifié. L'intérêt du *crowdsourcing* est donc d'économiser le temps précieux des spécialistes en remplaçant chacun d'eux par des milliers de joueurs dont les comptages vont être combinés pour aboutir à un diagnostic collectif de haute précision. D'après Miguel Luengo-Oroz, 3 % du temps que la population mondiale passe à jouer à des jeux vidéos seraient plus que suffisants pour diagnostiquer tous les cas de paludisme dans le monde ⁸. Le concepteur évoque également la possibilité d'utiliser les données générées par le jeu pour automatiser le diagnostic et minimiser le temps requis pour un comptage, et de développer un système d'acquisition d'images de microscopie par téléphone mobile pour permettre le transfert rapide des images vers la plateforme de télédiagnostic. Les résultats ont montré que combiner 22 parties de non-experts assure une

Jeux médicaux

précision dans le dénombrement des parasites de l'ordre de 99 %. On obtient le même résultat en combinant 13 parties de joueurs entraînés.

Pour les besoins du test et de la validation de la méthode, les images proposées aux joueurs ont déjà été caractérisées par des professionnels. Mais par la suite il est prévu d'ajouter des images non encore diagnostiquées pour augmenter la base de données.

Pour les concepteurs, les résultats obtenus valident l'approche ludique pour le comptage de parasites sur frottis sanguin par *crowdsourcing*, en montrant que des non-experts deviennent rapidement capables d'identifier l'agent de la malaria avec une précision similaire à celle de techniciens formés. L'expérience illustre le potentiel de l'approche ludique non seulement pour le diagnostic et la quantification de la malaria, mais plus généralement pour l'analyse d'images à des fins de télédiagnostic dans le contexte des grands problèmes de santé publique.

Compétence humaine requise/éducation du joueur : le jeu nécessite seulement un sens de l'observation et de la rapidité, puisque un élément de ludification est de limiter le temps d'observation à une minute. Comme pour *Malaria Training Game*, aucune connaissance préalable n'est requise et le joueur se familiarise juste avec les différents aspects que peut prendre le *plasmodium* sur un frottis sanguin.

Challenge/ludification : le jeu est immédiatement accessible : la page d'accueil tient lieu de tutoriel en montrant trois exemples de parasites tels qu'ils se présentent sur le type de préparation microscopique qui va être à explorer, et deux éléments qui ne sont pas des parasites, le tout avec un design emprunté à l'univers du *cartoon*, et une musique censée être entraînante. En cliquant sur *play*, une image est chargée et on dispose alors d'une minute pour la parcourir et trouver tous les parasites présents. Les éléments rendant la tâche de reconnaissance ludique sont divers : on ne voit pas l'image entière, on la déplace à la souris à travers un cercle noir qui simule un objectif de microscope. Au fur et à mesure des *clics*, un compteur indique au joueur combien il lui reste de parasites à découvrir en même temps qu'un compte à rebours défile, pimementant la recherche. Le résultat est donné sous forme de score, le nombre de parasites détectés est indiqué ainsi que le nombre d'erreurs. Si on n'a pas découvert suffisamment de parasites, la mention *vous avez découvert moins de 90 % des parasites* s'affiche et invite à rejouer. Dans le cas contraire, on passe au *niveau 2*. A la fin de chaque partie le joueur est invité à s'identifier et à envoyer son score, et à consulter le top 20 des *chasseurs de malaria*. Le jeu est également disponible sur iPhone et iPad.

2.6. Worm Watch Lab

Type de tâche concernée : nanotâche

But/présentation/généralités : *Worm Watch Lab*⁶ est un jeu accessible via le portail *Zooniverse*. C'est un projet du Medical Research Council (Medical Research Foundation) lancé en juillet 2013. Le but annoncé est une meilleure compréhension du fonctionnement du cerveau et de l'influence des gènes sur le comportement. Le principe est que si un gène est impliqué dans un comportement visible, alors une mutation affectant ce gène devrait se traduire par une modification notable du comportement. Le modèle d'étude est le nématode *Caenorhabditis elegans*, classiquement utilisé en biologie, en raison de la rapidité de son cycle de développement, du fait que son génome est entièrement séquencé, et que 40 % de ses gènes ont des équivalents chez l'homme. C'est donc l'occasion de tester facilement l'impact de certaines mutations sur le comportement. Le jeu consiste à visionner de courtes séquences vidéo montrant un nématode, et de cliquer sur la touche z chaque fois qu'on le voit pondre. En effet, chez le nématode, la ponte est le résultat d'un processus métabolique, contrôlé par la sérotonine, neurotransmetteur également impliqué dans la régulation de l'humeur chez l'homme. Ainsi certains antidépresseurs, qui stimulent la production de sérotonine chez l'homme agissent également sur le nématode en stimulant sa ponte. La mise en évidence, chez le ver, d'une augmentation ou d'une diminution de la ponte par rapport à la normale peut traduire un dysfonctionnement neuronal ou musculaire d'origine génétique et être informatif sur les mécanismes de la dépression chez l'homme.

Intérêt du crowdsourcing pour le problème à résoudre : dépister les vers qui présentent des anomalies dans le rythme de ponte est extrêmement long, car cela suppose d'en observer des milliers. Comme par ailleurs l'équipe de recherche du MRC a mis au point un système d'observation microscopique couplé à un ordinateur qui enregistre les mouvements du ver, il est extrêmement simple de visualiser et de signaler chaque ponte. Comme ça ne nécessite aucune compétence ni connaissance particulière, l'idée de la faire faire par des adeptes de la *science citoyenne* est à l'origine de *Worm Watch Lab*.

Compétence humaine requise : quasiment aucune autre que... de la patience et un *tempérament contemplatif*.

Challenge/ludification : tout est fait pour rendre attrayante et ludique l'idée de... regarder des vers se tortiller mollement et expulser un œuf de temps en temps. Mais comme ce n'est pas spécialement excitant, les concepteurs tentent de susciter l'adhésion par la motivation « aidez les chercheurs », et par l'humour « partez à la chasse ! ». L'interface est bleu vif, le ver est représenté façon *cartoon*... En cliquant sur *Start*, une vidéo de 30 secondes démarre. Pour chaque vidéo figure une

⁶*Worm Watch Lab* est accessible à l'adresse <http://www.wormwatchlab.org/>.

Jeux médicaux

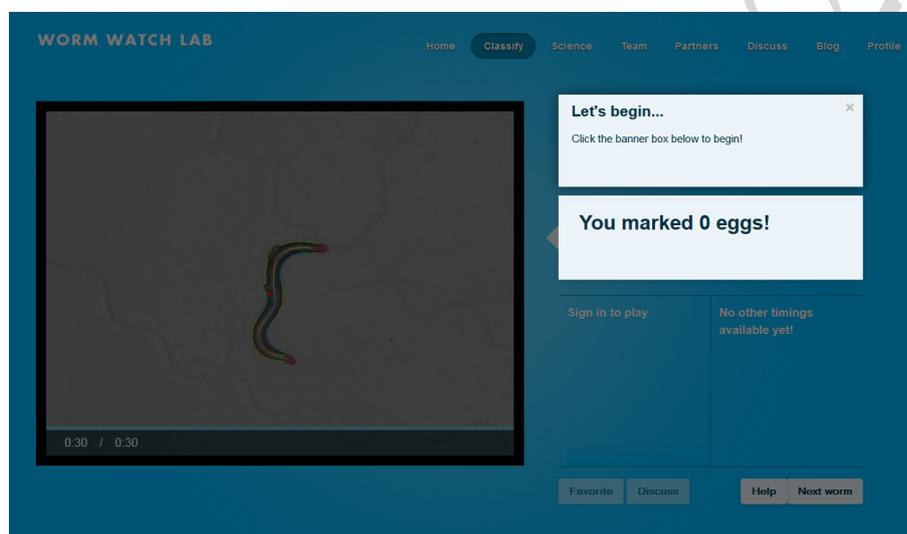


Figure 2.5.* Worm Watch Lab. Sur la gauche de la figure, l'écran permet de visualiser un ver en mouvement durant 30 secondes. On clique sur la touche z chaque fois qu'on le voit expulser un œuf. A droite s'affiche le résultat d'une séquence d'observation, ainsi que le nombre d'œufs observés par un autre joueur, si la séquence a déjà été visionnée.

information brève sur le type de mutation affectant le ver que l'on observe, si toutefois elle est connue. Le score est égal au nombre d'œufs détectés multiplié par 10. Il est remis à zéro à chaque nouvelle vidéo, mais le profil enregistre les scores cumulés des différentes vidéos observées, qui sont également mémorisées dans le profil sous forme d'icônes. Un bonus de 200 points est accordé lorsqu'on est le premier à regarder une vidéo. Par contre, si un autre joueur a également visionné la vidéo et signalé des pontes, son résultat est affiché, mais ne semble pas influencer sur le score en cas de différence d'appréciation sur le nombre d'œufs pondus.

Interactivité : à part l'affichage du nombre d'œufs vus par l'autre joueur (quand la vidéo a déjà été visionnée), on ne trouve aucune information sur les autres participants, aucun classement, alors que l'on s'attendrait, comme c'est le cas dans beaucoup de GWAP, à une validation par agrément, c'est-à-dire à marquer des points lorsqu'on déclare le même nombre d'œufs que tout autre joueur ayant visionné la même séquence. On peut néanmoins communiquer avec les autres et avec l'équipe de recherche dans un espace de discussion, et notamment poser des questions sur une séquence en particulier, soit à l'issue de l'observation en cliquant sur *discuss*, soit depuis son *profil*, en cliquant sur l'icône d'une vidéo.

2.7. Play to Cure : Genes in Space

Type de tâche concernée : microtâche

jeu disponible seulement pour certains modèles de smartphones – non testé.

Audience/popularité : lancé en février 2014, le jeu a permis, en à peine un mois et demi, de faire le travail d'analyse qui aurait pris 6 mois à un chercheur. Cela représente plus de 53 000 heures de jeu⁹.

But : *Play to Cure* est un jeu lancé par le CRUK (Cancer Research UK), le Centre de Recherche sur le Cancer du Royaume-Uni, également à l'origine de *Cell Slider*, une plate-forme de *crowdsourcing* dévolue au typage de cellules tumorales. Le but du jeu est de faire analyser par les joueurs les données issues de puces à ADN afin de mettre en évidence des zones altérées du génome humain. Aux commandes d'un engin spatial, le joueur parcourt le cosmos à la recherche de *substance alpha*. Le trajet emprunté par le vaisseau spatial en fonction de la contrainte imposée par le scénario permettra aux chercheurs de localiser des parties du génome pouvant correspondre à des signatures de tumeurs. Pouvoir ainsi cibler les régions du génome à analyser préférentiellement pour identifier des régions potentiellement impliquées dans tel ou tel type de cancer représente un gain de temps précieux¹⁰.

Intérêt du crowdsourcing : l'intérêt principal réside dans la myriadisation [Sagot *et al.*, 2011] de la tâche : l'analyse des puces à ADN est extrêmement longue et la répartir sur des milliers de personnes représente un gain de temps considérable.

Challenge/ludification : bien que nous n'ayons pas pu le tester (parce que nous n'avons pas de smartphone compatible), ce jeu semble être un des rares *vrais jeux*, d'après les vidéos de présentation et la forte médiatisation de son lancement : la tâche austère, monotone et ingrate est totalement dissimulée derrière un scénario de science-fiction servi par un design en 3D plutôt convaincant. Le joueur, aux commandes d'un vaisseau spatial, est investi d'une mission : collecter la précieuse *substance Alpha* en passant par les zones les plus denses du décor, tout en évitant et en détruisant des astéroïdes. Tous les éléments typiques du jeu vidéo sont présents : niveaux à passer, obstacles à éviter, tirs, possibilité d'augmenter la puissance de son vaisseau, etc.

Pour booster la participation, *Play to Cure* mise à fond sur la ludification, mais sans négliger pour autant de jouer sur la corde sensible : « *Be a science hero! Play our new free mobile game to help our scientists to beat cancer sooner* ».

Notes

⁷<http://www.hfsp.org/frontier-science/frontier-science-matters/crowdsourcing-design-swarming-nanoparticles-cancer-applica>.

Jeux médicaux



Figure 2.6.* Play to Cure : Genes in Space

⁸<https://www.ashoka.org/fellow/miguel-angel-luengo-oroz>.

⁹<http://www.gamesindustry.biz/articles/2014-03-17-play-to-cure-has-already-analysed-6-months-worth-of-cancer-data>.

¹⁰<http://scienceblog.cancerresearchuk.org/2013/03/01/can-the-power-of-the-public-help-personalise-cancer-treatment/>.

Jeux médicaux

Version auteurs, dec. 2014

CHAPITRE 3

Jeux pour le traitement automatique du langage

Dès la révolution informatique dans la seconde moitié du XX^e siècle, en particulier avec l'apparition des microprocesseurs dans les années 1970, les scientifiques ont exploité la supériorité de la machine sur l'homme concernant la rapidité des calculs. En ce qui concerne le traitement des données, et en particulier les données lexicales, les logiciels développés alors ont clairement montré des facilités d'accès aux informations, qui n'étaient même pas imaginables avec les traitements traditionnels. S'est alors posé le problème de l'acquisition de ces données, problème d'autant plus crucial que les ressources nécessaires étaient de plus en plus volumineuses. L'avènement d'Internet, qui a permis l'accès à des bases d'information de plus en plus importantes, a accru ce phénomène de façon exponentielle. Ce n'est qu'au début du XXI^e siècle, avec la démocratisation d'Internet, que des méthodes collaboratives ont vu le jour. L'un des exemples les plus célèbres est la constitution de l'encyclopédie collaborative Wikipédia. Les collaborateurs ont pleinement conscience qu'ils contribuent, de façon *sérieuse*, à la création d'une ressource de grande ampleur. Les GWAP constituent une autre manière d'aborder la création collaborative de ressources, où les internautes, tout en *jouant*, contribuent, généralement sans même s'en douter, à la constitution d'une ressource.

Dans ce chapitre, après avoir explicité la nécessité des ressources lexicales en traitement automatique du langage (TAL), nous présenterons la délicate question de leur acquisition, puis nous décrirons les principaux jeux dédiés au TAL, toujours actifs au moment de l'écriture de cet ouvrage.

Compte tenu de l'évolution dans ce domaine, notre liste n'est pas nécessairement exhaustive, et nous n'avons retenu que les jeux les plus pérennes et ceux ayant abouti à la constitution de ressources de qualité.

3.1. Pourquoi des ressources lexicales ?

Les ressources lexicales constituent la principale source d'informations linguistiques pour le traitement automatique du langage, le TAL, au premier rang duquel on trouve la compréhension automatique de texte. Qu'il s'agisse d'une compréhension *fine* (par exemple : quelle est la signification de cette phrase ? voire, cette phrase a-t-elle un sens ?) ou d'une compréhension plus *grossière* (par exemple : sur quel domaine porte ce texte ?), il est indispensable d'avoir à sa disposition des ressources lexicales pour pouvoir répondre à ces questions. D'une façon globale ces ressources portent à la fois sur les faits de langue (*matou* est synonyme de *chat*) et sur la réalité du monde (un *chat* a une *queue*). De manière plus précise, parmi les tâches nécessitant des ressources lexicales, nous pouvons citer la détermination de la proximité sémantique et de la similarité, l'extraction d'informations lexicales, l'annotation sémantique de documents, l'élaboration de paraphrases, la désambiguïsation sémantique, la traduction automatique, etc.

Ces ressources lexicales sont principalement conçues sous la forme de dictionnaires donnant le(s) sens des mots ou alternativement (et plus rarement) sous forme de réseaux lexicaux explicitant les relations possibles entre les mots. Ces deux formes ne sont pas indépendantes ; nous verrons au chapitre 5, avec la présentation de *JeuxDeMots* que le réseau lexical JDM peut prendre l'apparence d'un dictionnaire, baptisé *Diko*.

Quelles que soient leurs formes, ce sont nécessairement des ressources relativement volumineuses : un dictionnaire traditionnel contient plusieurs dizaines de milliers d'entrées, Wikipédia comporte environ 1,5 million d'articles dans sa version française, le réseau lexical de *JeuxDeMots* compte plus de 400 000 termes et 12 millions de relations. Signalons toutefois que, pour une communication « courante », un locuteur n'a besoin que d'environ 5 000 mots.

Il nous paraît nécessaire d'attirer ici l'attention du lecteur sur l'accessibilité des ressources. De plus en plus de ressources sont librement disponibles, leurs propriétaires en autorisant l'exploitation, voire la transformation ou l'enrichissement. Un nombre croissant d'auteurs estime en effet que le libre accès est le meilleur moyen dont disposent les développeurs de ressources lexicales pour relever avec succès les quatre défis principaux que sont :

- la visibilité : une ressource qui n'est pas visible, ne pourra pas être trouvée par ses utilisateurs potentiels, elle ne sera donc pas utilisée ;
- la disponibilité immédiate : c'est une condition indispensable pour qu'une ressource soit utilisée ;

- la disponibilité pérenne : des ressources librement accessibles ont plus de chances d'être utilisées sur le long terme, y compris lorsque les développeurs ont cessé d'y travailler ;
- l'évolutivité : une ressource qui porte sur la langue, doit pouvoir évoluer avec elle.

3.2. Les GWAP pour le traitement automatique du langage

Dans le cadre du traitement automatique du langage, les premiers GWAP sont apparus vers le milieu de la première décennie du XXI^e siècle. C'est donc un domaine relativement récent, actuellement en forte expansion, non encore *stabilisé* : pour preuve, un certain nombre de GWAP ont eu une durée de vie réduite, de quelques mois à un tout petit nombre d'années, et la ressource produite n'a pas été maintenue ; de plus, si elle n'a pas été mise en libre accès, elle n'est généralement plus disponible.

3.2.1. Le problème de l'acquisition des ressources lexicales

Les méthodes d'acquisition des ressources lexicales ont grandement évolué ces dernières décennies. Actuellement, on peut les classer en cinq grandes catégories :

- Acquisition automatique, à partir de corpus : de très grands corpus sont disponibles actuellement, grâce au Web, et des traitements automatiques sont technologiquement possibles dans des temps non prohibitifs. Le résultat peut cependant être bruité, d'où la nécessité d'un filtrage ne conservant que les mots significatifs. Le résultat peut également être biaisé, soit par le corpus lui-même (les corpus généralistes, tels des articles de journaux, font abstraction de connaissances élémentaires ou évidentes pour un humain, et les corpus encyclopédiques ne mentionnent généralement pas des informations relatives à l'actualité), soit par la méthode d'extraction (qui peut, par exemple, favoriser les cooccurrences ou mal interpréter certaines assertions) ; de plus, toute l'information n'est pas nécessairement disponible dans les textes (information intuitive, sous-entendue). Ce processus nécessite une vérification manuelle, au moins sur un échantillon de la ressource acquise.

- Acquisition par le biais de contributions, uniques ou faiblement redondantes, rémunérées ou non. Ces contributions sont souvent le fruit d'un travail d'experts, long et coûteux, et le résultat est souvent figé, difficilement évolutif (l'un des exemples les plus caricaturaux est le dictionnaire de l'Académie française). Dans ce cas, la vérification, manuelle par experts, n'apparaît généralement pas indispensable. De plus, une fois la ressource *officialisée* et mise à disposition, toute correction peut s'avérer difficile (par exemple, la plupart des dictionnaires « papier » n'ont qu'une édition annuelle).

- Acquisition au moyen de micro-contributions, fortement redondantes, d'un grand nombre de contributeurs : ce sont des systèmes d'acquisition par *myriadisation d'un travail parcellisé* (ou MTP)

Jeux et TAL

dont on trouvera une analyse critique dans [Sagot *et al.*, 2011]. Les contributeurs, souvent appelés *turkers*, sont généralement (faiblement) rémunérés, comme dans le service Amazon Mechanical Turk. Ce type de méthode repose sur le fait que de nombreux internautes sont prêts à collaborer à des micro-tâches, le plus souvent au prix d'une rétribution symbolique, parfois même bénévolement. Le grand nombre de contributeurs impliqués va permettre la constitution de ressources de grande taille. Malheureusement, l'expérience montre que la qualité n'est pas toujours au rendez-vous : certains biais peuvent être introduits par le découpage en micro-tâches et il est difficile d'obtenir une qualité acceptable sur des tâches complexes. Ce mode d'acquisition de données n'est pas propre au domaine du TAL : le site d'Amazon Web Services ¹ présente une version d'Amazon Mechanical Turk comme « une place de marché pour les travaux faisant appel à l'intelligence humaine » ; les exemples donnés concernent les traitements photo/vidéo (par exemple : la classification d'objets dans des images), la vérification/nettoyage de données (par exemple : la détection de doublons dans des catalogues), la collecte d'informations diverses (par exemple : répondre à des enquêtes), etc.

- Acquisition grâce à des jeux contributifs, les GWAP donc, où les joueurs contribuent à l'acquisition d'une ressource, parfois sans en être conscients : il s'agit d'une ressource correspondant à une connaissance générale commune, élaborée par un grand nombre de joueurs, d'où un risque d'erreurs, involontairement introduites ou non, qu'il est impératif de minimiser. C'est un processus relativement rapide, peu coûteux et évolutif. Cette approche est motivée par l'observation qu'un groupe d'individus agissant de manière non concertée obtient généralement de meilleurs résultats qu'un individu isolé, même si celui-ci est un expert du domaine. Le nombre de joueurs potentiels sera ici d'autant plus grand que les connaissances nécessaires pour jouer (les connaissances fondamentales de la langue naturelle) sont partagées par le plus grand nombre. De plus, les systèmes collaboratifs tels que Wikipédia ont clairement montré l'engouement des internautes pour ces projets. Il faudra toutefois vérifier la validité de la ressource ainsi produite, mais cette vérification peut également être effectuée au moyen de GWAP (par exemple, *Totaki* dans le projet *JeuxDeMots*). La frontière entre la MTP non rémunérée et les GWAP, parfois quelque peu floue, se situe uniquement au niveau de la dimension ludique de ces derniers, ces deux méthodes d'acquisition de ressources reposant sur le phénomène de myriadisation (ou peuplonomie).

- Acquisition ou enrichissement des ressources lexicales dans une langue à partir de ressources déjà disponibles dans d'autres langues. Divers procédés ont été développés ; ils reposent principalement sur l'hypothèse selon laquelle les concepts et les relations entre concepts sont indépendants de la langue, hypothèse qui peut introduire un biais dans la ressource obtenue.

¹<http://aws.amazon.com/fr>.

3.2.2. Ressources lexicales actuellement disponibles

La création des ressources lexicales est étroitement liée à l'histoire de l'écriture. Elle remonte au moins au III^e millénaire avant JC, avec les collections tabulaires mésopotamiennes qui constituent les premières ressources *dictionnairiques* monolingues. Parallèlement, sont apparues des ressources bilingues, voire multilingues, permettant la traduction et la transmission d'informations entre les peuples. L'un des exemples les plus célèbres est la fameuse pierre de Rosette qui permit à Champollion de déchiffrer les hiéroglyphes.

Deux révolutions technologiques majeures, l'imprimerie au milieu du XV^e siècle puis l'informatique dans la seconde moitié du XX^e, vont avoir une influence prépondérante sur la façon de concevoir, élaborer et diffuser les ressources lexicales. Les premiers dictionnaires dignes de ce nom vont voir le jour au début du XVII^e siècle, la première édition du *Dictionnaire de l'Académie* est publiée en 1694, et les dictionnaires « papier » tels que nous les connaissons aujourd'hui, exhaustifs et avec des explications détaillées, apparaissent au cours du XIX^e siècle. Un ouvrage important est publié en 1852 : il s'agit du *Thesaurus of English Words and Phrases* de Roget ; il est basé sur la notion de concepts et permet ainsi l'accès aux mots à partir du sens. A la fin du XX^e siècle, la plupart des grands dictionnaires sont informatisés, ce qui en facilite l'accès et la mise à jour ; leur contenu évolue avec l'introduction de données multimédia ; les méthodes d'acquisition des ressources se diversifient, avec l'apparition de méthodes (semi-)automatiques et des approches collaboratives (dont les GWAP qui nous intéressent ici).

Les ressources lexicales électroniques apparaissent au tout début du XXI^e siècle. Ce sont des ressources destinées à être utilisées principalement par la machine. Les bases de données lexicales fournissent une description structurée et explicite des entrées (les unités lexicales). Les informations fournies sont d'ordre phonologique, catégorielle (par exemple : la catégorie grammaticale), quantitative (par exemple : la fréquence d'apparition dans un corpus), lexicale (par exemple : synonymie), etc. Parallèlement à ces bases de données lexicales (dont les noms se terminent souvent par *-lex*), apparaissent les réseaux lexico-sémantiques (dont les noms se terminent souvent par *-net*). Chaque nœud de réseau est une unité lexicale (terme ou concept) et les liens reliant ces nœuds expriment des relations de caractère lexical (synonymie, antonymie, ...), ontologique (hyperonymie, ...) ou d'un autre type (association libre, sentiment associé, ...). Citons ces principaux réseaux :

- WordNet [Miller *et al.*, 1990] et [Fellbaum, 1998] : parfois appelé le Princeton Wordnet, c'est le plus célèbre, le début de sa construction remonte à 1990 et il comporte plus de 100 000 nœuds, les synsets, assimilables à des concepts regroupant chacun un certain nombre de termes.

A sa suite, et parfois s'appuyant sur sa structure, d'autres réseaux sont apparus, parmi lesquels :

- EuroWordNet [Vossen, 1998], puis BalkaNet, SloWNet, AsianWordNet : des équivalents de WordNet pour les langues européennes ou asiatiques ;

Jeux et TAL

- BabelNet [Navigli et Ponzetto, 2010] qui comporte 3 millions de concepts et implémente des connaissances lexicales et encyclopédiques ;

- HowNet [Dong et Dong, 2006] dont chacun des 1 500 concepts est caractérisé avec un ensemble d'attributs et de propriétés (relations lexicales, rôles sémantiques, ...);

- WOLF (WordNet Libre du Français) [Sagot et Fiser, 2008], librement disponible, élaboré par des techniques automatiques ;

Certains projets ont émergé indépendamment de WordNet :

- Cyc (projet débuté en 1984) [Lenat, 1995] : réseau sémantique de sens commun basé sur la logique, en accès libre qui compte 500 000 concepts ;

- *JeuxDeMots* [Lafourcade, 2007] réseau lexico-sémantique, libre d'accès, en construction permanente depuis 2007, comporte plus de 420 000 termes reliés par 12 millions de relations ; il est construit collaborativement au moyen de GWAP (voir le chapitre 5 présentant le projet JDM).

La plupart de ces réseaux peuvent faire l'objet de certaines critiques. Un certain nombre d'entre eux ne sont pas libres d'accès, ce qui en limite sérieusement l'utilisation, comme nous l'avons vu au début de ce chapitre. De plus, les relations entre termes sont généralement non typées (par exemple : la relation entre chat et animal y paraît analogue à celle entre chat et souris) et rarement pondérées (les relations fréquentes sont alors implémentées de manière analogue à des relations inusuelles ou rares). Enfin, la détermination des sens est soit inexistante, soit simplifiée ou trop complexe, à l'instar des synsets dans WordNet, qui correspondent en définitive assez peu à l'usage des locuteurs. Parmi les réseaux cités, seul *JeuxDeMots*, libre d'accès, est constitué de relations typées et pondérées qui permettent de déterminer, à l'aide des joueurs, des usages différents pour un même terme (polysémie). Ces différents sens (ou raffinements) sont également pondérés, ce qui permet de mettre en relief le sens le plus communément admis.

Vers ce qui serait désirable

Sans être utopiste, ni verser dans la science-fiction, interpréter des textes comme pourrait le faire un humain, nécessite idéalement de disposer d'un réseau lexical et sémantique constitué de relations typées et pondérées entre termes. Ce réseau devrait à la fois comporter des termes généralistes, représentant une *culture générale commune* afin de pouvoir traiter ce qui relève de la vie courante, mais également des termes de spécialités, étiquetés par cette (ou ces) spécialité(s), afin de pouvoir interpréter des textes issus de domaines spécifiques. Un tel réseau devrait comporter au minimum plusieurs centaines de milliers de termes reliés par un nombre de relations de l'ordre de quelques millions. (A titre de simple comparaison numérique, le cerveau humain compte approximativement cent milliards de neurones.) Ces relations proviendraient à la fois de connaissances du

monde (par exemple, *un chat mange de la pâtée*, ou *chasse des souris*) et de connaissances lexicales (par exemple, *chat* a pour synonyme *greffier*).

La pondération d'une relation relativement à une autre est difficile à quantifier. A quoi correspond-elle ? A la *force intrinsèque* d'une relation (difficile à définir), ou à sa fréquence (sur quelle base, ou par rapport à quel corpus ?). De plus, la force d'une relation peut dépendre du contexte. Par exemple, on conçoit aisément que, d'une manière générale, la relation chat-souris ait un poids supérieur à la relation chat-allergie, alors que dans le domaine médical la seconde relation est prépondérante. Dans JDM, la solution simple (simpliste ?) qui a été adoptée établit la pondération par comptage (une fonction logarithmique du comptage serait probablement plus proche de la perception humaine). Il serait possible également d'envisager une pondération des termes eux-mêmes, éventuellement en fonction de leur fréquence (dans quel ensemble de corpus ?). Ici aussi, cette pondération serait dépendante du domaine concerné, généraliste ou de spécialité.

Un tel réseau rend possible la détermination des différents sens d'un terme, et ce de manière peut être plus fine que celle rencontrée dans les dictionnaires traditionnels, ce qui peut conduire à une détermination des usages : l'arborescence des sens est à un voire deux niveaux dans les dictionnaires traditionnels, alors qu'elle comporte généralement plusieurs niveaux dans le cas du réseau lexical JDM. Mais attention, là aussi, la détermination des sens peut dépendre du contexte ; par exemple le terme « IVG » signifie généralement « Interruption Volontaire de Grossesse », y compris dans le domaine médical, à l'exception du sous-domaine de la cardiologie où il prend la signification d'« Insuffisance Ventriculaire Gauche ».

En résumé, une ressource lexicale qui serait désirable est un réseau lexico-sémantique :

- en perpétuelle construction-évolution (pour tenir compte des évolutions de la langue, des découvertes, de l'actualité, etc.) ;
- où les termes seraient pondérés par leur fréquence d'usage, mais également tagués par leur(s) domaine(s) d'application ;
- où les termes ayant des sens ou des usages multiples seraient *raffinés* en autant de sous-termes (par exemple *chat* seraient relié à *chat>animal* et *chat>discussion*) ;
- avec des relations typées, pondérées, orientées ; en plus des relations *classiques* (par exemple, synonymie, antonymie, hiérarchie, etc.), il pourrait comporter des relations *moins standard* (par exemple, des relations relatives aux affects ou au jugement, divers rôles sémantiques, des relations de typicalité pour les lieux, les valeurs temporelles, etc.) ;
- qui pourrait comporter également des objets agglomérés avec des relations entre un couple ou triplet et d'autres termes (par exemple, la phrase « dans les contes, les ogres mangent les petits enfants avec avidité » constitue un exemple où un couple de termes est relié à d'autres termes :

Jeux et TAL

manger [sujet] ogre associé par une relation de manière à *avec avidité*, relation de patient (objet) à *enfant* ou *petit enfant* et relation de lieu typique à *conte*).

En fait, la lecture des chapitres suivants et notamment de celui consacré au projet *JeuxDeMots* montre bien que le réseau lexico-sémantique visé par ce dernier est un exemple de ressource qui cherche à satisfaire ces caractéristiques.

3.2.3. Intérêts des GWAP en TAL

Les avantages de l'utilisation des GWAP pour l'acquisition de ressources lexicales sont multiples :

- les traitements en TAL nécessitent des ressources volumineuses, probablement plusieurs centaines de milliers (voire millions) de termes et plusieurs dizaines (voire centaines) de millions de relations entre ces termes pour représenter non seulement les ressources généralistes de connaissances communes, mais également les ressources spécialisées si l'on veut traiter des domaines spécifiques : le recours aux GWAP permettrait la constitution d'un réseau *universel* généraliste et multi-spécialisé, quasi impossible à réaliser avec les autres techniques actuelles ; le problème de l'acquisition de ressources (voir section 3.2.1), commun à d'autres domaines (voir les chapitres 1, 2 et 4), est donc d'autant plus crucial en TAL que les ressources à collecter sont particulièrement volumineuses ;

- contrairement à d'autres méthodes d'acquisition, les GWAP permettent de prendre en compte l'aspect dynamique et évolutif de la langue et des connaissances : effectivement, on assiste constamment à l'apparition de mots nouveaux (par exemple, *virus H5N1*, *peuplonomie*) et de sens nouveaux pour des mots déjà existants (par exemple, *virus*, *tablette*) ;

- de plus, tout n'est pas écrit dans les textes, en particulier les informations évidentes pour un humain (par exemple, avez-vous déjà rencontré une phrase vous déconseillant l'utilisation de votre smartphone sous la douche ?) ; il est donc nécessaire de collecter des ressources d'usage (et non normatives), lesquelles sont généralement implicites dans les textes.

- enfin, les GWAP permettent la constitution de ressources par consensus ; les langues naturelles ne faisant pas partie des sciences exactes, la ressource acquise sera d'autant plus fiable qu'elle résulte de consensus entre plusieurs personnes, plutôt que d'émaner d'une seule.

Dans ce qui suit, nous présentons les principaux GWAP dédiés de près ou de loin au traitement automatique du langage et à l'acquisition de ressources lexico-sémantiques.

3.3. PhraseDetectives

Type de tâche concernée : microtâche, même si la tâche à accomplir peut être relativement complexe.

Lancement : décembre 2008

La première version de *PhraseDetectives*² a été mise en ligne en décembre 2008. Ce jeu a été conçu par trois universitaires britanniques de l'Université d'Essex : Jon Chamberlain, Massimo Poesio et Udo Kruschwitz [Chamberlain *et al.*, 2008].

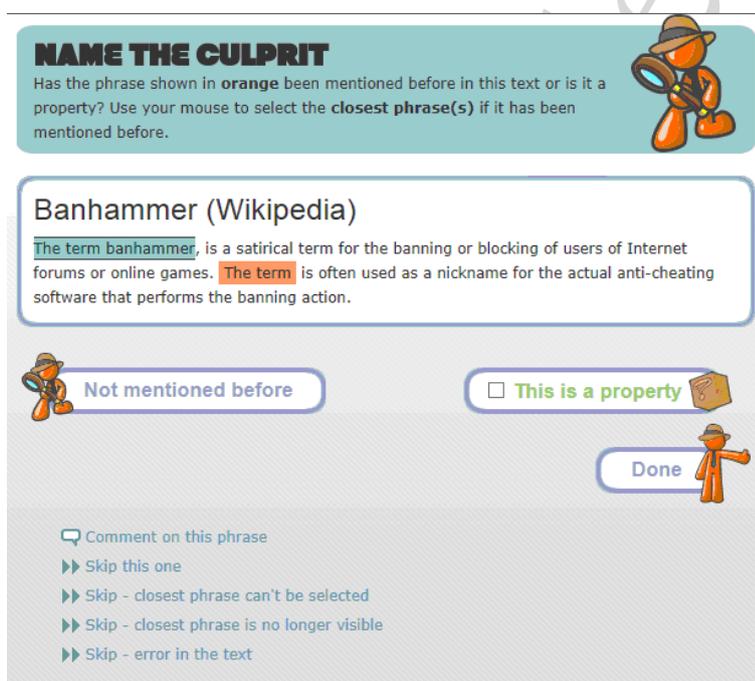
But : annotation de liens anaphoriques, ce qui à la base, représente une tâche relativement complexe.

Figure 3.1.* PhraseDetectives : « Name the culprit ». Il est demandé au joueur quel terme (s'il existe dans le texte présenté) constitue l'antécédent de l'anaphore surlignée en orangé. Lorsque le joueur passe sa souris sur la phrase, les termes ou groupes de termes se surlignent en bleu. Le joueur a également la possibilité d'affirmer qu'il ne s'agit pas d'une anaphore, en cliquant sur « Not mentioned before » ou sur « This is a property ».

²*PhraseDetectives* est accessible à l'adresse <http://www.phrasedetectives.org>.

Jeux et TAL

De quoi s'agit-il ? En grammaire, une anaphore (aussi appelée anaphore grammaticale, par opposition à l'anaphore en rhétorique) est un mot qui, dans un énoncé, assure une reprise sémantique d'un précédent segment appelé antécédent. Il existe donc, entre l'antécédent et l'anaphore, une relation : l'anaphore est l'élément représentant, et l'antécédent est l'élément représenté. L'anaphore permet ainsi d'éviter les répétitions. Le premier exemple donné dans *PhraseDetectives* est le suivant : « *Sherlink Holmes went to the shop. He got some tobacco for his pipe* ». Le mot « *He* » est surligné et il est indiqué qu'il fait bien évidemment référence à « *Sherlink Holmes* ».



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.

Banhammer (Wikipedia)

The term **banhammer**, is a satirical term for the banning or blocking of users of Internet forums or online games. **The term** is often used as a nickname for the actual anti-cheating software that performs the banning action.

Not mentioned before

This is a property

Done

- Comment on this phrase
- Skip this one
- Skip - closest phrase can't be selected
- Skip - closest phrase is no longer visible
- Skip - error in the text

Figure 3.2.* *PhraseDetectives* : « *Name the culprit* ». Ecran montrant la réponse du joueur, surlignée en bleu. Il ne lui reste plus qu'à cliquer sur « *Done* »

PhraseDetectives comporte ainsi une phase d'entraînement où le joueur novice apprend la tâche qu'il va devoir réaliser, au moyen d'exemples réalisés par des experts. Ensuite, le joueur doit annoter un texte ; si plusieurs joueurs saisissent des annotations différentes pour le même texte, celui-ci est présenté à un plus grand nombre de joueurs pour validation. Contrairement à la plupart des GWAP, ici, le joueur a pleinement conscience qu'il contribue à une tâche *sérieuse* : il n'est guère amusant, hormis pour « les amoureux de la littérature, de la grammaire et du langage » (comme il est dit sur la page d'accueil de *PhraseDetectives*), de réaliser des annotations sur des liens anaphoriques.

Le principal aspect *jeu* de ce GWAP réside dans sa présentation, avec un petit personnage virtuel assimilé à un détective et un système de points permettant de réaliser un classement entre les joueurs. De plus, une version pour Facebook a été développée, permettant un accès *via* les réseaux sociaux.



Figure 3.3.* PhraseDetectives : « *Detectives Conference* ». Il est demandé au joueur de valider, ou non, la proposition faite précédemment par un autre joueur. Si le joueur désapprouve la proposition, il aura la possibilité d'indiquer quel terme, selon lui, constitue l'antécédent de l'anaphore, de manière analogue au mode « *Name the Culprit* »

De manière plus détaillée, *PhraseDetectives* comporte deux modes/niveaux de jeu : un mode d'annotation, puis un mode de validation. Avant même de pouvoir accéder au premier mode, le joueur novice suit un apprentissage au cours duquel lui sont présentées des phrases pour lesquelles les réponses correctes sont connues ; en cas de réponses invalides de la part du joueur, des explications lui sont fournies pour l'aider à comprendre ses erreurs. Quand un joueur a donné suffisamment de bonnes réponses, il peut accéder au premier mode dans lequel ses réponses seront enregistrées. Dans ce premier mode, intitulé « *Name the Culprit* » (Nommez le Coupable), un texte est présenté au joueur dans un mode d'annotation directe, où une anaphore est surlignée. Un exemple en est donné à la figure 3.1. Le joueur doit alors indiquer quel(s) mot(s) du texte constitue(nt) l'antécédent de cette anaphore. La réponse du joueur sur l'exemple de la figure 3.1 est présentée à la figure 3.2. Il peut arriver que le joueur estime que le terme surligné n'est pas une anaphore, auquel cas il peut le mentionner en cliquant sur « *Not mentioned before* » ou « *This is a property* » : dans ce dernier

cas, il indiquera de quel terme c'est une propriété (afin d'aider les joueurs, des exemples de ces différents cas sont donnés sur le site). Si les avis des joueurs diffèrent sur une même phrase, alors leurs différentes interprétations sont présentées à un plus grand nombre de joueurs dans le deuxième mode, dénommé « *Detectives Conference* » (la Conférence des Détectives). Un exemple d'écran de ce deuxième mode est reproduit à la figure 3.3. Dans ce mode de validation, les joueurs expriment leur accord, ou leur désaccord, avec les interprétations précédemment faites.

3.4. PlayCoref

Type de tâche concernée : microtâche, même si la tâche à accomplir peut être relativement complexe.

Lancement : 2009

*PlayCoref*³ a été développé à l'Université Charles, à Prague, par Barbora Hladka et Jiri Mirovsky [Hladka *et al.*, 2009].

But : annotation de coréférences dans des données textuelles.

On parle de coréférence lorsque plusieurs termes dans un même texte font référence à la même entité ; en linguistique, une coréférence est reliée à son antécédent. Par exemple, dans la phrase « le petit chat est mort, il est tombé du toit, c'est comme ça, il a glissé sur j'sais pas quoi », les deux pronoms personnels « il » font référence au terme « le petit chat » ; ces trois termes forment une chaîne de coréférences. Lors d'une partie, le joueur devra, en un temps imparti, créer le plus de liens possible, en connectant entre elles toutes les coréférences qu'il identifie dans le plus grand nombre de phrases possible d'un même texte.

Une partie de jeu débute par l'affichage des premières phrases d'un texte dans lequel les termes pouvant être des coréférences apparaissent en caractères gras ; le joueur a la possibilité de sélectionner uniquement ces termes afin de constituer, selon lui, une chaîne de coréférences. La figure 3.4 reproduit un exemple d'écran après que le joueur ait indiqué cinq paires de coréférences, créant ainsi cinq liens. Dès qu'il pense avoir fini, le joueur peut demander l'affichage de la phrase suivante en cliquant sur « *Next* », et ainsi ajouter des coréférences, éventuellement entre des termes n'appartenant pas à la même phrase. Le joueur a également la possibilité de supprimer des liens qu'il vient de créer. La partie s'arrête soit au bout du temps imparti (cinq minutes), soit lorsque le joueur pense avoir traité tout le texte et décide de terminer la partie en cliquant sur « *Finish* ». Un score est alors affiché, récompensant le joueur pour son travail, et l'invitant à poursuivre avec une nouvelle partie. Une partie de *PlayCoref* peut être jouée par un joueur en solitaire, ou entre deux joueurs. Dans le cas d'une partie jouée en solitaire, le score est calculé par comparaison entre les propositions du joueur

³*PlayCoref* est accessible à l'adresse <http://www.lgame.cz>.

Jeux et TAL

et celles réalisées manuellement ou par une procédure automatique. Dans le cas d'une partie entre deux joueurs, la comparaison entre leurs propositions est intégrée dans le calcul du score.

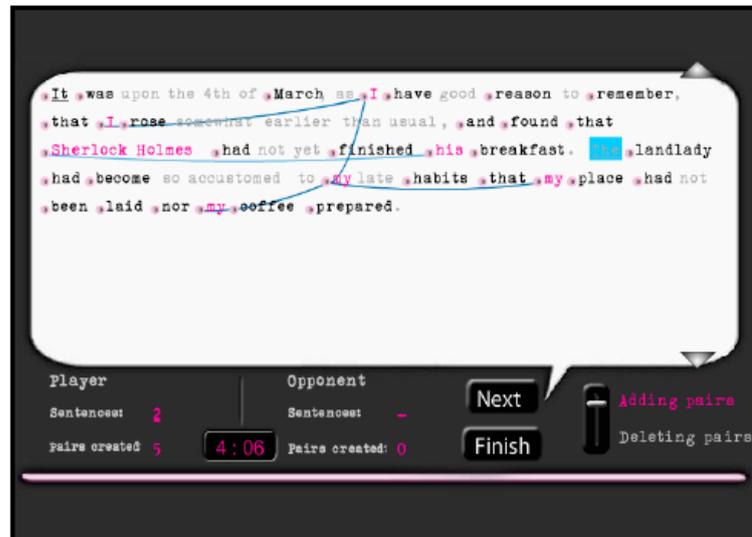


Figure 3.4.* PlayCoref. Exemple d'écran montrant les liens de coréférences proposés par le joueur

L'on notera une certaine analogie, dans le but recherché, entre *PlayCoref* et *PhraseDetectives* (vu à la section précédente). *PhraseDetectives* demande l'antécédent précédent, alors que *PlayCoref* demande toutes les coréférences ; dans *PhraseDetectives*, il est possible de fournir d'autres informations (telles que « Propriétés », par exemple).

L'analyse des premiers tests auprès des joueurs, réalisée par les auteurs de ce GWAP, montre des résultats très encourageants ; en particulier, certaines parties à deux joueurs ont permis de détecter un plus grand nombre de coréférences qu'une annotation manuelle (étaient-elles toutes valables ?).

Deux autres jeux sont accessibles depuis le portail de *PlayCoref* :

- *Shannon game* : une phrase avec des mots manquants est affichée ; le joueur a droit à plusieurs essais pour les deviner ;

- *Place the space* : le joueur doit insérer des caractères *espace* dans une phrase où tous les mots sont accolés.

Ces deux jeux sont-ils réellement des GWAP ? Nous ne le pensons pas.

3.5. Verbosity

Type de tâche concernée : microtâche, même si une partie de jeu est une forme de dialogue entre deux joueurs.

Lancement : 2005

Ce jeu a été développé par Luis Von Ahn (l'un des pionniers des GWAP), Mihir Kedia et Manuel Blum de l'Université Carnegie Mellon [von Ahn *et al.*, 2006].

But : collecter des *faits de sens commun*.

Un fait de sens commun est une assertion vraie qui est connue de la très grande majorité des humains (par exemple : « le feu brûle », ou « l'eau mouille »). Une telle base d'assertions est nécessaire afin de pouvoir programmer des systèmes *intelligents*. Le nombre de faits de sens commun (estimé à plusieurs centaines de millions, selon les auteurs) est tellement important que les méthodes classiques de collecte se sont toujours révélées inefficaces, d'où l'idée de recourir à un GWAP. Celui-ci se joue à deux joueurs en simultané : un joueur *descripteur* doit faire deviner un mot secret, déterminé aléatoirement par le système, à un joueur *devineur* en lui fournissant des indices ; les rôles sont inversés à chaque partie. L'interface utilisateur propose des zones de saisie de texte typé sémantiquement que le joueur *descripteur* est invité à renseigner ; le joueur *devineur* voit donc s'afficher les indices typés et peut faire des suggestions. Le premier joueur ne peut saisir comme indices que des termes connus du système. Le typage sémantique des indices repose sur des phrases à compléter (par exemple : « le terme cible est une sorte de... », ou « le terme cible est utilisé pour... », ou encore « le terme cible possède... »). La figure 3.5 reproduit un exemple de partie pour un joueur *devineur*, alors que la figure 3.6 présente un exemple de partie pour un joueur *descripteur*.

Un des problèmes majeurs est qu'il n'y a aucun contrôle sur les indices fournis par un joueur, en principe typés, mais ce typage n'est pas nécessairement respecté ; en pratique, le jeu se résume souvent à donner une série de termes indices librement associés au terme cible, et il s'amorce souvent une forme de discussion spontanée entre les deux joueurs. Qu'en est-il de la fiabilité des informations ainsi sauvegardées ? Difficile à dire, car les données collectées ne semblent pas libres d'accès. Signalons que dans JDM, si le typage n'est pas respecté, l'intersection entre les propositions des joueurs risque d'être vide.

Il semblerait que le site www.gwap.com à partir duquel il était possible d'accéder à *Verbosity* ne soit plus accessible (depuis peu) ; dans le blog de gwap.com, il est écrit que ses auteurs ont migré vers d'autres projets.

Jeux et TAL

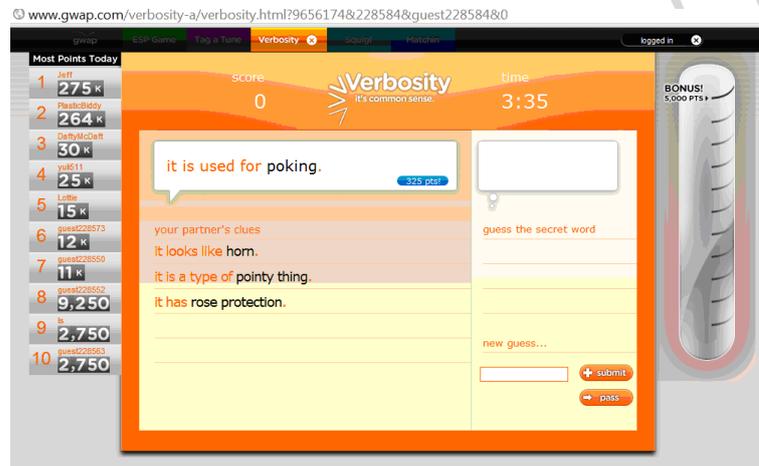


Figure 3.5.* Verbose : joueur devineur. Ici, le joueur doit deviner le mot secret à partir des indices proposés

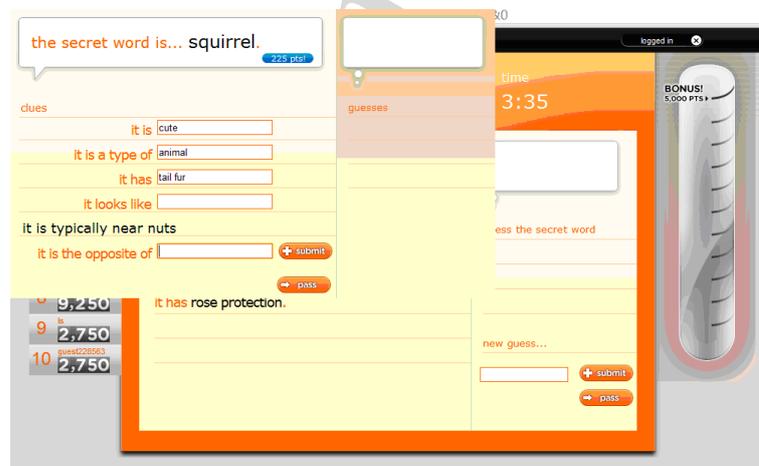


Figure 3.6.* Verbose : joueur descripteur. Ici, le joueur doit faire deviner le mot à partir des indices proposés. Les indices proposés ici suivent les consignes de types, mais aucune vérification n'est faite à ce sujet en temps réel. De plus, il est parfois difficile de renseigner certains indices (par exemple, ici : quel est le contraire de « écureuil » ?).

3.6. JeuxDeMots

Type de tâche concernée : microtâche pour la plupart des jeux du projet *JeuxDeMots*⁴. Pour le jeu principal, une partie dure généralement une minute, voire plus. On peut alors considérer qu'il s'agit d'une succession de microtâches.

Lancement : juillet 2007. Sa toute première présentation a été faite par [Lafourcade, 2007].

JeuxDeMots sera détaillé au chapitre 5 : il comporte différents jeux et contre-jeux, ces derniers permettant à la fois d'offrir de la diversité aux joueurs, et de vérifier les données collectées *via* le jeu principal. L'évaluation de la ressource acquise est également réalisée au moyen de *Totaki*, un GWAP où le joueur peut parfois *se sentir plus fort que la machine*. De plus, un petit nombre de joueurs a émis le désir de s'impliquer plus, en participant de façon plus *précise* à la construction du réseau : ces derniers ont donc la possibilité de *contribuer via* l'interface *Diko*, c'est-à-dire de renseigner directement, pour chaque terme du réseau, la ou les relation(s) de leur choix. Ces contributions seront validées manuellement par l'administration du jeu.

But : constitution d'un réseau lexical de termes reliés par des associations typées, pondérées, taguées. L'une des conséquences est la construction *automatique* d'un dictionnaire.

Le principal jeu du projet, appelé lui-même *JeuxDeMots*, met en lice deux joueurs, de manière anonyme et asynchrone. Pour un même terme cible, aléatoirement sélectionné dans une base de termes, et une même consigne, ces deux joueurs fournissent en un temps limité les réponses qui leur semblent correspondre (par exemple : « quelles sont les idées associées au terme *chat* ? » ou « quelles sont les parties de *voiture* ? »). Seules les réponses communes aux deux joueurs sont mémorisées. Ces réponses déterminent des relations entre le terme cible et les termes réponses. Les réponses des joueurs peuvent introduire des mots qui ne figuraient pas dans la base de termes, et ainsi enrichir celle-ci. La grande force de *JeuxDeMots* réside dans l'originalité du concept : parmi la multitude des jeux dits *de mots* ou *de lettres* disponibles sur Internet, il n'en existe aucun qui mobilise ce genre de compétence, à la fois pointue (pour être bon, il faut avoir un vocabulaire étendu et/ou être rapide et organisé pour faire un maximum de réponses dans le temps imparti) et universellement partagée (il s'agit de fournir des connaissances générales véhiculées par un support commun : la langue).

En bref, c'est un jeu accessible à tous, où l'on peut progresser et s'améliorer rapidement en observant et en imitant les stratégies des meilleurs joueurs. Les concepteurs du projet s'efforcent d'entretenir l'engouement suscité par la découverte du jeu en ajoutant régulièrement soit des nouveaux jeux, soit des variantes dans le jeu principal, de façon à éviter la lassitude en proposant un maximum de diversité. Depuis son lancement en 2007, plusieurs milliers de joueurs se sont connec-

⁴*JeuxDeMots* est accessible à l'adresse <http://www.jeuxdemots.org>.

tés, participant au total à plus de 1,4 million de parties pour le seul jeu principal. Le résultat est la constitution, à partir d'une base brute de 150 000 termes isolés, d'un réseau lexico-sémantique, qui compte aujourd'hui (octobre 2014) plus de 400 000 termes reliés par 12 millions de relations.

3.7. Zombilingo

Type de tâche concernée : microtâche

Lancement : 2014

*Zombilingo*⁵, [Fort *et al.*, 2014], est l'un des derniers GWAP apparu sur le Net, développé par Karèn Fort et Bruno Guillaume de l'Université de Lorraine.

But : annotation de corpus en syntaxe de dépendances.

Chaque phrase ajoutée dans la base est préannotée (sujet, verbe, compléments...) automatiquement à l'aide d'analyseurs syntaxiques. Il est ensuite demandé aux joueurs d'indiquer leur annotation pour cette phrase. Si un nombre jugé suffisant de joueurs donnent un avis contraire à la préannotation automatique, l'annotation de la phrase considérée est alors modifiée. Comme annoter une phrase complètement peut être une tâche relativement complexe, *Zombilingo* décompose la tâche globale en une série de tâches plus élémentaires afin de ne pas décourager le joueur. Ce dernier choisit un type de tâche (annotation du sujet, de l'objet, de l'attribut de l'objet ...), suit la formation correspondante sur un corpus de référence, et peut ensuite commencer à jouer.



Figure 3.7.* Zombilingo : annotation du sujet. Le joueur doit indiquer quel mot de la phrase est, selon lui, le sujet du verbe affiché en vert

⁵*Zombilingo* est accessible à l'adresse <http://zombilingo.loria.fr>.

Jeux et TAL

L'interface du jeu, le monde des zombies, a été particulièrement étudiée pour attirer les joueurs et les inciter à jouer, afin de produire suffisamment d'annotations. Un exemple d'écran, concernant l'annotation du sujet d'une phrase, est reproduit à la figure 3.7. Un deuxième exemple, concernant l'annotation de l'objet d'un verbe, est reproduit à la figure 3.8. Les données produites sont librement accessibles à partir du site.

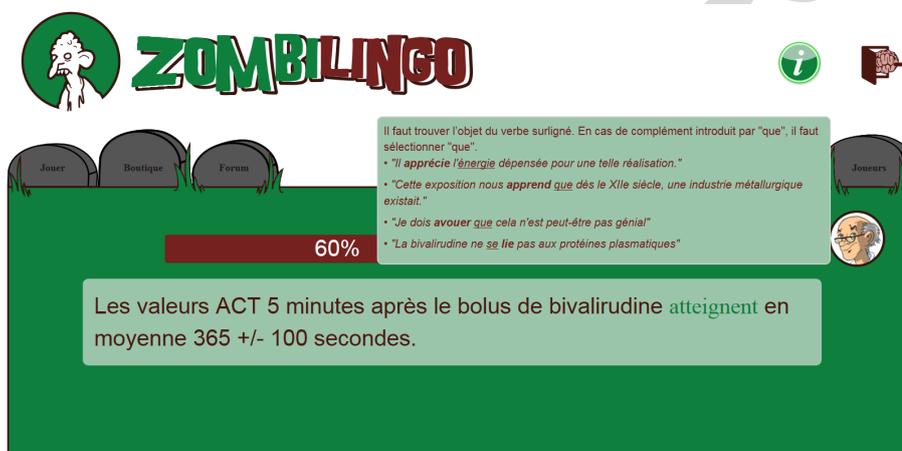


Figure 3.8.* Zombilingo : annotation du complément d'objet. Le joueur doit annoter le complément d'objet du verbe

Ce jeu s'adresse clairement à des gens qui gardent un bon souvenir de la grammaire à l'école, et pour qui l'analyse logique est un casse-tête divertissant. On peut néanmoins regretter que le libellé des tâches à accomplir soit quelque peu opaque : « P obj agt », « dep » ou encore « Affixe » ne renseignent absolument pas sur ce qu'il faut faire. Du coup, le joueur peut être tenté de cliquer au hasard dans la première phrase proposée, afin que la correction immédiate qui s'ensuit vienne éclaircir un peu les choses. Peut-être est-ce voulu, peut-être veut-on que le joueur découvre par lui-même, en tâtonnant et en se trompant, ce qu'on attend de lui... ou est-ce un *défaut de jeunesse* de ce GWAP, encore en phase de lancement ; parions que dans ce cas il sera très vite corrigé. Mais pour les gens qui ont des connaissances en grammaire, il est un peu agaçant de devoir se tromper pour découvrir à quelle notion grammaticale doit répondre l'annotation requise.

3.8. Infection

Type de tâche concernée : microtâche

Lancement : 2014

*Infection*⁶ est un jeu relativement hors-normes puisqu'il s'agit d'un Vidéo-GWAP dans la mesure où l'utilisateur ne fournit pas une réponse textuelle, mais se comporte comme dans un jeu d'action, en tirant sur des personnages fictifs (simplifiés, reconnaissons-le). Ce jeu a été développé par Daniele Vannella et Roberto Navigli de l'Université de Rome [Vannella *et al.*, 2014].

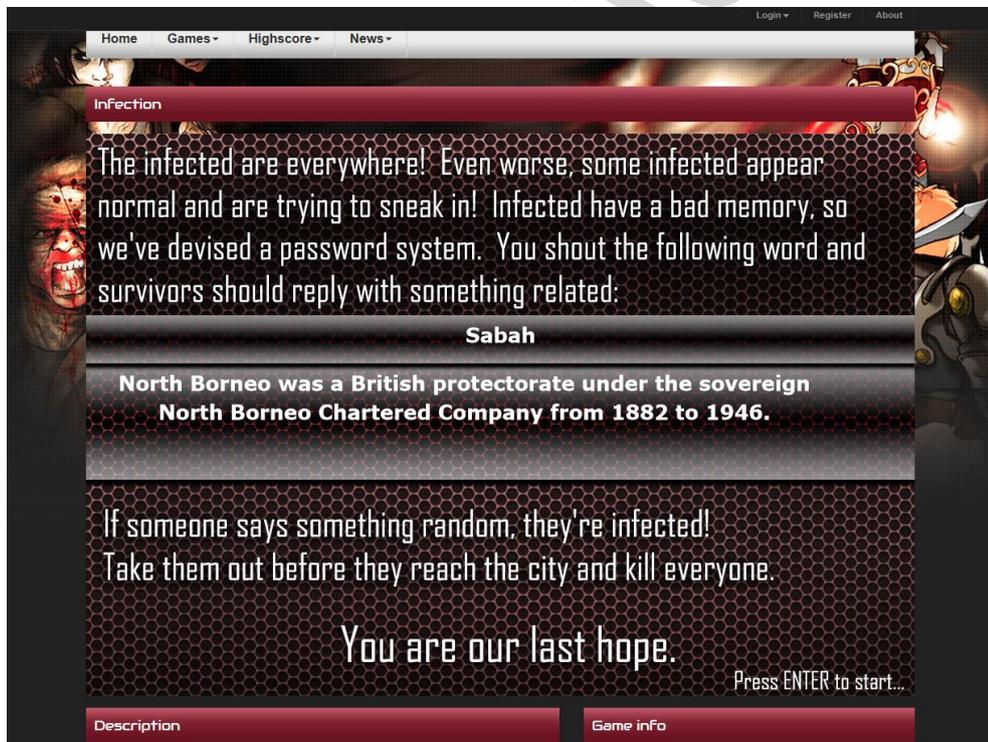


Figure 3.9.* Infection : la consigne

But : Validation d'associations entre concepts.

Dans le scénario d'*Infection*, l'humanité est attaquée par un virus : une partie de la population est encore saine, une autre partie est infectée. Les humains se dirigent tous vers une ville *a priori* non encore infectée. Le joueur doit stopper les humains infectés pour éviter la propagation de l'épidémie,

⁶*Infection* est accessible à l'adresse <http://knowledgeforge.org/>.

mais laisser passer ceux qui sont sains. Comme tous les humains se ressemblent, le jeu se base sur un système de question-réponse-mot de passe pour les distinguer : les réponses des humains sains sont cohérentes avec la question, celles des humains infectés ne le sont pas (par exemple, si le mot de passe est « médecine », un humain sain pourra répondre « radiologie » ou « médicament », alors qu'un humain infecté donnera une réponse sans rapport avec ce thème).

Le joueur a donc pour mission de laisser passer les humains qui, selon lui, répondent par un terme en rapport avec le *thème de passe*, et de stopper tous les autres. Le jeu est divisé en plusieurs niveaux de difficulté. Les joueurs obtiennent des points en fonction de leur capacité à sauver l'humanité ; un tableau général affichant les scores des joueurs permet de se situer et accroît la motivation. En modifiant la question mot de passe, *Infection* permet d'obtenir des relations d'association libre, de synonymie ou d'antonymie. Les joueurs sont supposés jouer honnêtement, mais un mécanisme simple contrôlant les réponses des joueurs a été mis en place afin de détecter les malveillances flagrantes.



Figure 3.10.* Infection : la partie lancée

Soulignons ici que [Jurgens et Navigli, 2014] développent une approche originale, dans la mesure où, bien que ces jeux concernent l'annotation ou la désambiguïsation lexicale, aucune action de nature textuelle n'est requise de la part du joueur, le mode de jeu se rapprochant davantage de celui des jeux de type *shoot 'em up*.

3.9. Wordrobe

Type de tâche concernée : microtâche

Lancement : septembre 2012

*Wordrobe*⁷ a été développé par Johan Bos et son équipe de l'Université de Groningen, aux Pays-Bas [Venhuizen *et al.*, 2013].

But : enrichir la *Groningen Meaning Bank* (GMB), un corpus sémantiquement annoté gratuitement accessible.

A l'origine, *Wordrobe* comportait un seul jeu, *Sense*, conçu pour recueillir des données de désambiguïsation lexicale afin de contribuer à la construction de la GMB, un outil de référence précieux pour le traitement automatique du langage. Le principe est celui du QCM : une phrase comportant un mot en caractères gras est affichée, et le joueur doit choisir dans une liste de propositions le sens approprié du mot en question. Une partie comporte une dizaine de phrases à annoter ainsi, et le nombre de points gagnés est fonction du degré de similarité des réponses avec celles des autres joueurs pour les mêmes phrases. Comme les parties sont mémorisées, le score varie même quand on ne joue pas, puisque les parties jouées sont confrontées à celles des autres. Un curseur permet de miser sur ses réponses ce qui pimente le jeu, puisque une mise importante est à double-tranchant : le gain potentiel est proportionnel à la mise, mais la perte potentielle l'est aussi. On peut tenter de flamber. . .

Wordrobe s'est rapidement enrichi de nouveaux jeux, et représente aujourd'hui une collection de sept jeux d'annotations diverses, tous basés sur le principe décrit ci-dessus. Ainsi, dans la phrase proposée, il peut s'agir de choisir le bon antécédent pour un pronom (*Pointers*), ou d'identifier le *type d'entité* que désigne un nom propre (*Names*), ou d'assigner un statut d'*agent*, de *cause* ou d'*instrument* à un terme en relation avec un verbe d'action (*Roles*), ou encore de dire si le terme indiqué désigne un humain, un animal, un objet concret, une entité abstraite, etc. (*Animals*)

Comme la difficulté est variable, l'ensemble des tâches proposées peut satisfaire tous les publics, « des linguistes amateurs aux gourous du langage » comme dit le slogan, à l'accueil. Comme beaucoup de jeux à but lexico-sémantique, il faut quand même être un amoureux éperdu de la langue et des subtilités du vocabulaire pour devenir un joueur très productif et assidu. Malgré la relative variété des tâches, le principe du jeu est toujours le même, et les éléments ludiques (pari, classement, récompenses diverses) ont du mal à rendre excitante à long terme cette activité d'annotation.

⁷*Wordrobe* est accessible à l'adresse <http://www.wordrobe.org>.

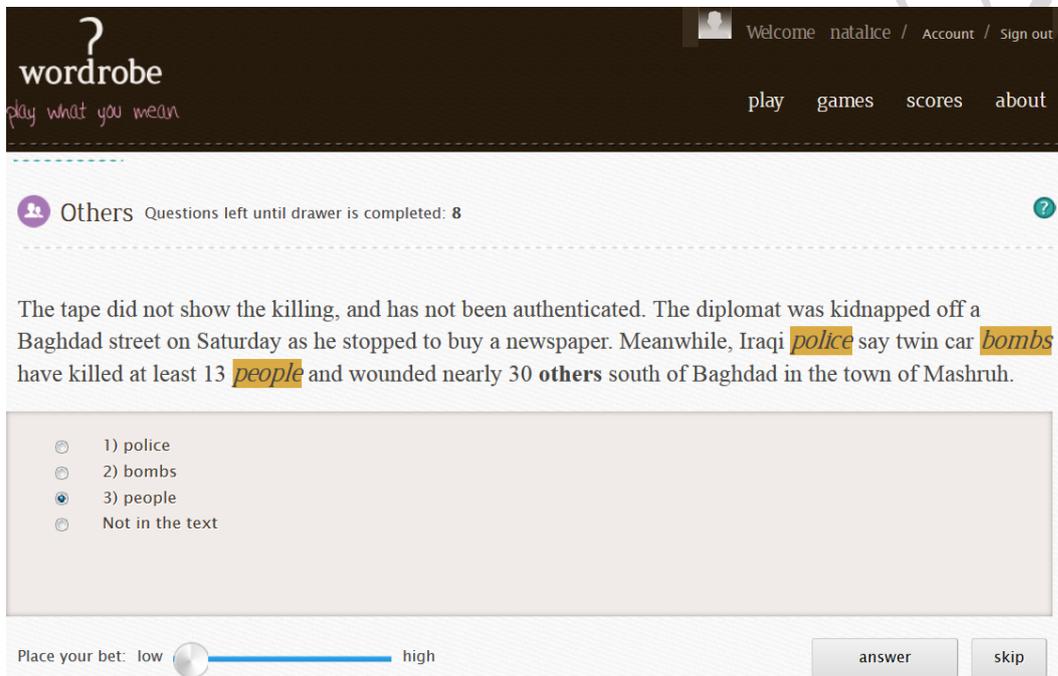


Figure 3.11.* Wordrobe : une partie de Others, un des sept jeux d'annotation. Dans la phrase affichée, il faut dire à quel mot surligné en jaune correspond le mot en caractères gras **others**, en cochant l'item correspondant dans la liste proposée dessous. Le curseur en bas à gauche permet de miser sur sa réponse, et de maximiser son gain (mais en prenant le risque d'obtenir l'effet inverse, si la réponse donnée n'est pas celle choisie par la majorité des autres joueurs).

3.10. D'autres GWAP dédiés au TAL

Ci-après, nous présentons très brièvement un certain nombre d'autres GWAP dédiés au TAL, qui représentaient d'intéressantes stratégies, mais qui malheureusement, ne semblent plus accessibles, ce qui pose le problème de la pérennité de l'approche GWAP. Le caractère éphémère de certains de ces jeux vient du fait qu'il s'agit souvent de prototypes, et que leur durée de vie est celle d'un projet et de son financement, ou encore du doctorat dont ils sont le sujet.

3.10.1. Open Mind Word Expert

Type de tâche concernée : microtâche

Lancement : 2002

But : c'est un système d'apprentissage dont le but est la collecte du marquage (tagging) des sens de mots polysémiques dans un corpus.

Open Mind Word Expert [Mihalcea et Chklovski, 2003] a été développé au début des années 2000 par Timothy Chklovski du MIT et Rada Mihalcea de l'Université du Texas. Dans ce jeu, plusieurs phrases comportant un même mot polysémique sont proposées au joueur, ainsi que les différents sens de ce mot extraits de WordNet. Pour chaque occurrence affichée du mot polysémique, le joueur doit indiquer quel sens parmi ceux présentés lui paraît le plus adéquat ; il a également la possibilité de cliquer sur « *unclear* » ou sur « *none of the above* ».

De par son principe d'acquisition de ressources, *Open Mind Word Expert* est l'un des ancêtres des GWAP ; mais, compte tenu de son interface (qui n'a rien de ludique), peut-il être réellement considéré comme un jeu ? Pour preuve, ses auteurs en le présentant parlent d'utilisateurs ou de contributeurs, mais jamais de joueurs. Nous parlerions plutôt de système d'acquisition par myriadiation d'un travail parcellisé (MTP).

3.10.2. 1001 Paraphrases

Type de tâche concernée : microtâche

Lancement : 2005. *1001 Paraphrases* est l'un des tout premiers GWAP, développé par Timothy Chklovski [Chklovski, 2005].

But : collecte de données permettant aux systèmes de traduction automatique de mieux reconnaître les différentes variantes de paraphrase.

Dans une partie, il est demandé aux joueurs d'indiquer des paraphrases pour une expression qui leur est donnée (par exemple : « ceci peut vous aider »). Si le joueur indique une paraphrase déjà précédemment fournie par un autre joueur, il gagne un certain nombre de points ; sinon, sa paraphrase est ajoutée à celle déjà proposées, en attente des futures parties. C'est probablement le premier GWAP dédié au TAL utilisant la méthode de validation par paires de joueurs (paires ou plus, dans le cas de *1001 Paraphrases*).

3.10.3. Categorilla/Categodzilla

Type de tâche concernée : microtâche

Lancement : 2008

But : collecte d'associations entre des termes et leurs caractéristiques.

Ce GWAP, qui comporte deux variantes, a été développé par l'équipe de David Vickrey à l'Université de Stanford [Vickrey *et al.*, 2008]. Dans *Categorilla*, il est demandé aux joueurs de fournir des termes (mots ou expressions) correspondant au critère fourni (par exemple : « indiquez des objets

qui volent ») et commençant par une lettre donnée (par exemple : la lettre « B » ; et dans ce cas-là les réponses peuvent être « *bird* » ou « *Boeing* »). Chaque partie a lieu entre deux joueurs, qui reçoivent des points lorsque leurs réponses concordent (tout comme dans JDM). Les joueurs peuvent saisir autant de réponses qu'ils le souhaitent, et à la fin de chaque partie toutes les réponses sont affichées. Ce jeu a permis d'obtenir un nombre important de relations de type hyperonymie/hyponymie, ou fournissant les rôles sémantiques de patient/agent typique (relation unissant une action à son patient ou agent typique).

Categozilla est une version légèrement différente : la contrainte correspondant à la première lettre est soit supprimée, soit subdivisée entre lettres faciles et lettres difficiles.

3.10.4. FreeAssociation

Type de tâche concernée : microtâche

But : collecte d'associations entre termes.

Dans ce jeu, également développé par David Vickrey à Stanford, il est demandé aux joueurs de fournir des mots en relation avec un terme cible indiqué, sans pouvoir donner des termes *interdits* provenant d'une liste *taboue*. Le principe de ce GWAP ressemble à celui de JDM, en interdisant aux joueurs les mots les plus fortement reliés au terme cible, et donc en favorisant l'émergence de relations moins courantes, appartenant à la longue traîne⁸ (les nombreuses relations peu fréquentes).

3.10.5. Entity Discovery

Type de tâche concernée : microtâche

Lancement : 2010

But : collecte d'entités nommées.

Ce jeu a été développé par l'équipe de Nathan Green [Green *et al.*, 2010] à l'Université de Caroline du Nord. Le joueur doit indiquer les types des entités nommées qu'il détecte dans la phrase proposée ; il a le choix entre « Personne », « Lieu », « Organisation ». En pratique, après une phase d'apprentissage (où les réponses aux phrases proposées sont connues), le joueur se voit proposer des phrases pour lesquelles le système ne possède pas encore de réponse. Chaque partie se joue entre deux joueurs, anonymement, de manière synchrone ; cela permet d'indiquer à chaque joueur, à la fin de la partie, les concordances entre ses réponses et celles de l'autre joueur, qui déterminent le nombre de points obtenus.

Une variante de ce GWAP, appelée *Name That Entity*, se joue de manière asynchrone : le second joueur exerce alors une forme de vérification des réponses fournies par le premier.

⁸http://fr.wikipedia.org/wiki/Longue_tra%C3%Aene.

3.10.6. PhraTris

Type de tâche concernée : microtâche, même si la tâche demandée n'est pas toujours *simpliste*.

But : détermination de l'ordre des termes dans une phrase.

Il s'agit un jeu d'annotation syntaxique, développé par l'équipe de Giuseppe Attardi à l'Université de Pise [Attardi et the Galoap Team, 2010]. Dans ce jeu, les joueurs doivent réorganiser des phrases de manière logique, tout comme il faut réorganiser des briques dans le célèbre jeu Tetris.

Un certain nombre de GWAP ont brièvement vu le jour, mais méritent néanmoins d'être mentionnés.

- *Common Consensus*, développé par l'équipe d'Henry Lieberman du MIT, a été officiellement présenté pour la première fois en 2007 dans [Lieberman *et al.*, 2007]. Il est destiné à collecter les connaissances de sens commun sur les buts qui motivent les humains dans leurs actions de la vie courante, ainsi que les moyens mis en œuvre pour y parvenir. La principale raison ayant conduit à l'élaboration de ce jeu est la remarque exprimée par ses auteurs sur le fait que les ordinateurs ne possèdent pas un grand nombre d'informations de base sur le comportement et l'environnement des humains (par exemple : « les humains boivent quand ils ont soif », ou « le ciel est bleu pendant la journée »). Selon les auteurs, il est estimé que la quantité de ces éléments de connaissance se chiffrent en centaines de millions, voire en milliards, d'où l'idée de développer un GWAP pour les collecter.

Le principe de *Common Consensus* est analogue à celui d'un jeu télévisé (*Une famille en or* dans les pays francophones, *Family Feud* dans les pays anglo-saxons) où deux familles s'affrontent en répondant à des questions simples et ouvertes et où leurs réponses vont être comparées aux réponses de la communauté. Dans le jeu télévisé, les réponses de la communauté sont prédéfinies, élaborées par sondages ; dans le cas de *Common Consensus*, les réponses de la communauté sont définies dynamiquement en se basant sur les réponses des joueurs. Lors d'une partie de *Common Consensus*, un but est initialement aléatoirement sélectionné dans une base de buts préexistante (par exemple : « regarder un film »), puis les questions posées reposent sur une grille de questions relatives à ce but (par exemple : « pourquoi voudriez-vous... », ou « quels objets utilisez-vous pour... »). Certaines réponses à ces questions permettent de déterminer de nouvelles questions à poser dans de prochaines parties, construisant ainsi une hiérarchie parmi les buts. Lors d'une partie, limitée dans le temps, les joueurs peuvent saisir autant de réponses qu'ils le souhaitent. Chaque réponse saisie est comparée à l'ensemble des réponses déjà fournies par d'autres joueurs pour cette même question ; le joueur obtient un retour immédiat sous la forme d'une barre de taille proportionnelle au nombre de joueurs

ayant fourni cette même réponse. Un programme adéquat permet de comparer deux réponses afin de déterminer si elles sont sémantiquement similaires ou non.

Il semblerait que *Common Consensus* ne soit plus accessible.

- *Train Robots* est un jeu d'annotation, similaire à *PhraseDetectives*. Le joueur dispose de deux images d'une configuration de blocs colorés, l'une représentant une situation initiale et l'autre une situation finale. Un bras robotisé figure également à l'écran. Le joueur doit fournir, en langage naturel, la succession d'opérations qui permettra au bras robotisé d'obtenir la configuration finale à partir de la configuration initiale. Le jeu réalise ainsi la collecte de données lexicales permettant de programmer des systèmes robotiques *intelligents*.

- *OnToGalaxy* est un jeu développé par Markus Krause et son équipe à l'Université de Brème. Son objectif est l'acquisition de connaissances de sens commun sur les mots, en particulier afin de compléter des ontologies. L'interface de ce Vidéo-GWAP représente une scène de bataille spatiale ; le joueur, aux commandes d'un vaisseau spatial, doit tirer sur des objets, selon une consigne donnée. Celle-ci est donnée en termes de jeu, et non en termes d'acquisition de ressources (par exemple : « tirez sur tous les vaisseaux représentant des objets que l'on peut toucher » signifie « quels sont les objets qui ont la propriété d'être concrets ? »). Agissant ainsi, le joueur associe des objets cibles à des termes sources et des prédicats (par exemple : « dans l'image présentée, quels sont les objets plus gros qu'une voiture ? »), ou bien des objets possédant une propriété donnée (par exemple : « dans l'image présentée, quels sont les objets chauds ? »).

- *FACTory Game* : dans ce jeu, une affirmation est affichée ; l'utilisateur peut alors indiquer si, selon lui, elle est vraie, fausse, s'il ne sait pas ou si elle n'a aucun sens (par exemple : « des fils électriques sont placés dans des logements sans salle de séjour »). La réponse fournie par le joueur est comparée à la réponse majoritaire des autres joueurs ; le nombre de points qu'il obtient est fonction du résultat de cette comparaison.

D'autres jeux de lettres et de mots

De nombreux autres jeux de lettres et de mots existent sur le Web ; certains ressemblent à des GWAP car ils mémorisent les réponses des joueurs lorsqu'elles sont différentes de celles attendues par le système. Mais peu d'entre eux sont réellement des GWAP dans la mesure où il n'y a aucune vérification des données (qui peuvent être erronées dans ce cas) fournies par les joueurs, et donc aucun apprentissage valable n'est réalisé par le système : les parties jouées n'ont pas comme résultante la constitution d'une ressource de qualité.

Parmi ces jeux, nous pouvons citer par exemple :

Jeux et TAL

- *GuessWhat?* : il s'agit également d'un jeu de devinettes, relativement simple. Le joueur doit deviner un mot correspondant aux indices affichés à l'écran (par exemple : « animal à fourrure douce, avec des crocs pointus, des moustaches et habituellement une longue queue ») et dont le nombre de lettres est indiqué (par exemple : « 3 », pour un mot à deviner en langue anglaise). La réponse du joueur est enregistrée par le système si elle est différente de celle attendue. Le joueur a également la possibilité d'abandonner s'il ne trouve pas. Ce n'est pas un GWAP, aucune ressource réellement valide n'étant créée par les joueurs.

- *Find the link* : le logiciel affiche deux mots, ou plus, caractéristiques du mot à trouver ; le joueur a le choix entre plusieurs propositions. Contrairement à *JeuxDeMots*, c'est un jeu fermé, ce n'est pas le joueur qui fait les propositions d'associations, il n'y a aucun apprentissage par la machine.

Jeux et TAL

Version auteurs, dec. 2014

CHAPITRE 4

Jeux inclassables

Ce chapitre *fourre-tout*, évidemment non exhaustif, illustre la très grande diversité des domaines d'études et/ou de recherche susceptibles de donner lieu à des GWAP. La diversité s'accompagne d'une relative disparité quant à la popularité des jeux, qui reflète elle-même directement l'intérêt qu'ils suscitent auprès de la masse de contributeurs potentiels que représente le grand public. Notons que la corde sensible et la fibre citoyenne sont plus faciles à mobiliser par des projets dont les retombées sont d'ordre médical, que par l'incitation à aider les chercheurs en physique quantique, ou à participer à diverses tâches de classification plus ou moins fastidieuses. Pour compenser, dans les domaines où l'adhésion n'est pas liée aux affects, ni à l'empathie suscités par les grands enjeux de santé publique, les GWAP doivent motiver les contributeurs potentiels en *tirant sur d'autres ficelles*. Autrement dit, le plaisir procuré par le jeu doit être d'autant plus grand que la valorisation liée à la *grandeur* de la cause servie est modeste.

4.1. Beat the Bots

Type de tâche concernée : ni microtâche, ni macrotâche

But/présentation : le concepteur du jeu est la société ShareThinkLtd qui est à l'origine de *VouchSafe*¹, un système dédié à la lutte contre les robots spammeurs. *VouchSafe* se veut une alternative ludique au système des Captchas, agréable à utiliser, et d'une plus grande efficacité. Le principe est

¹*VouchSafe* est accessible à l'adresse <http://www.vouchsafe.com/>.

Jeux inclassables

le suivant : au lieu de saisir au clavier une série de lettres plus ou moins lisibles, l'utilisateur qui veut s'identifier en tant qu'humain va devoir résoudre une énigme visuelle simple. Les énigmes sont de deux types :

- trouver les deux objets ayant le plus de rapports entre eux : confronté à cinq ou six images d'objets usuels qui apparaissent simultanément à l'écran, il va devoir relier d'un trait l'objet qu'on lui désigne dans l'instruction textuelle à celui qui lui est le plus apparenté au niveau fonctionnel ou thématique ;

- trouver l'intrus : Parmi les cinq à six objets présentés, il faut désigner celui qui n'a aucun rapport avec les autres en l'entourant d'un trait.

*Beat the Bots*² est le premier prototype de jeu destiné d'une part à tester la technologie *VouchSafe*, et d'autre part à récolter des données afin d'instruire le logiciel sur la façon de raisonner des humains, afin qu'il soit plus efficace dans la discrimination humain/robot.

En outre, les concepteurs s'élèvent contre le recours au *crowdsourcing* classique pour effectuer des tâches complexes ou de grande envergure. C'est selon eux une méthode qui donne de bons résultats mais qui est critiquable d'un point de vue éthique. Ils préconisent plutôt l'approche ludique, qu'ils baptisent *crowdplay*¹.

Intérêt du crowdsourcing dans le problème à résoudre : la technologie *VouchSafe* exploite le fait que cerveau humain diffère radicalement d'une machine dans sa façon de percevoir une information visuelle et d'y associer du sens. Rien n'est aussi efficace que de faire répondre un humain... pour tester un système destiné à discriminer les réponses humaines et celles générées automatiquement par un logiciel. Les associations que font spontanément les joueurs entre les objets présentés sont utilisées pour aider ce dernier à concevoir de nouveaux challenges visuels, plus intuitifs et plus humains.

Compétence humaines requises : aucune... puisqu'il s'agit juste d'être humain... et de réagir avec spontanéité et rapidité, cette dernière qualité n'étant utile que pour l'aspect *jeu*.

Interactivité : comme *Fraxinus*, *Beat the Bots* est accessible via Facebook et utilise donc toute l'infrastructure du réseau social pour favoriser sa propre diffusion en incitant les joueurs à solliciter leurs amis. On joue contre la machine et non pas contre un autre humain, donc l'interactivité réside plutôt dans l'incitation à comparer son score avec celui de ses amis, et à s'exprimer sur le jeu en rejoignant la page Facebook dédiée.

Challenge/ludification : le jeu est amusant et on se laisse facilement aller à faire plusieurs parties de suite. La tâche, d'apparence simple, est pimentée par la limitation de temps : une partie comprend 15 groupes d'images et on dispose de 2 minutes pour traiter l'ensemble. (Il est possible de passer une série d'images moyennant une pénalisation (perte de points)). On n'a donc pas le temps d'hésiter,

²*Beat the Bots* est accessible à l'adresse <https://apps.facebook.com/beatthebots/>.

Jeux inclassables

ce qui favorise les réponses spontanées. A l'issue de la partie, divers paramètres de classement sont communiqués : le nombre de points, qui est fonction du nombre de réponses justes, mais aussi du temps mis pour répondre, le temps de réponse le plus rapide pour la série, le nombre de réponses justes, et la plus longue succession de réponses justes. Le but est faire un assez bon score pour entrer dans le *hall of fame*, qui affiche les 100 meilleurs joueurs. En outre, les joueurs du *top 5* et du *top 15* peuvent afficher des *trophées*, respectivement d'*or* et d'*argent*, sur leur *mur*.

Design : le design est simple, coloré, un peu *clinquant*, et s'accompagne d'une musique quelque peu lancinante, et d'effets sonores heureusement désactivables. La prise en main est immédiate, une partie ne dure que 2 minutes, il est donc possible d'en faire rapidement une en passant sans devoir s'immerger dans des règles et des choix stratégiques compliqués, ou en enchaîner plusieurs en fonction du temps dont on dispose. Un des gros avantages, (au niveau jouabilité, et par conséquent au niveau utilisation de la procédure de discrimination robot/humain) mis en avant par les concepteurs est le suivant : comme jouer ne nécessite rien d'autre que tracer une ligne ou faire un cercle, le jeu, comme le produit fini, sont particulièrement adaptés aux écrans tactiles des smartphones et tablettes.

4.2. Apetopia

Type de tâche concernée : ce jeu insolite est difficilement classable selon le critère micro-tâche/macrotâche.

But : *Apetopia*³ est un jeu lancé par l'Université de Berlin, et destiné à fournir des données sur la manière dont les nuances de couleur sont perçues par les gens, afin de modéliser de meilleurs paramètres colorimétriques.

Intérêt du crowdsourcing/compétences requises : l'intérêt est directement lié à la grande variabilité avec laquelle le cerveau humain discrimine les nuances de couleur, qui va permettre de constituer une large base de données. Aucune compétence cognitive particulière n'est requise, si ce n'est, pour l'aspect jeu, la dextérité et la rapidité nécessaires pour les jeux vidéo de type *running game*.

Challenge/ludification/Design : la ludification est très poussée, le joueur peut complètement oublier qu'il collabore à la production de données scientifiques. L'interface ressemble à celle de n'importe quel jeu vidéo : des successions de passages à deux portes défilent à l'écran, il faut, en naviguant avec les touches du clavier, franchir les portes dont la couleur se rapproche le plus de celle du ciel, en évitant divers obstacles (qui font perdre des *points de vie*) et en collectant autant de *pièces d'or* (bonus) que possible. La succession des portes s'accélère au fur et à mesure de la progression,

³*Apetopia* est accessible à l'adresse <http://colors.htw-berlin.de/>.



Figure 4.1.* Apetopia. L'écran d'accueil du jeu

rendant le défi de plus en plus difficile. Le jeu ne nécessite pas d'inscription, on choisit à la fin de communiquer son score, ou non, pour être inscrit dans un classement quotidien et global.

Commentaire : l'aspect jeu est parfaitement convaincant, que ce soit au point de vue décor, ergonomie, challenge. *Apetopia* est un jeu simple voire simpliste, accessible, tout public, qui ne nécessite aucune compétence particulière si ce n'est une dextérité manuelle que possèdent tous les adeptes de ce genre de jeu (pas nous, qui sommes d'une désespérante nullité). Il est difficile d'aller plus loin dans l'analyse car on ne trouve aucune explication détaillée sur le but scientifique, la pertinence des résultats obtenus par rapport au but recherché, la manière dont ils sont exploités. Dommage...Serait-ce un projet abandonné ?

4.3. Quantum Moves

Type de tâche concernée : plutôt macrotâche

But : le jeu a été lancé en version beta en 2012, via le portail *Science at home*⁴ mis en ligne par le département de physique et d'astronomie de l'Université d'Aarhus au Danemark. Entre le moment où nous l'avons testé et la rédaction de ces pages, il a été interrompu. Sa réactivation prochaine sous une version améliorée est annoncée sur le portail. Le jeu est basé sur la simulation des

⁴*Science at home* est accessible à l'adresse <http://www.scienceathome.org/>.

Jeux inclassables

mouvements atomiques que l'on obtient en laboratoire au moyen d'un faisceau laser. Dans le jeu, ce sont les mouvements de la souris qui simulent le faisceau laser et l'objectif est de déplacer un atome dans la zone cible de la manière optimale et dans le plus bref délai. Les solutions des joueurs ayant réussi le meilleur score sont comparées avec celles issues d'algorithmes de calcul, et testées et optimisées en laboratoire pour identifier lesquelles peuvent être utilisées dans la résolution de problèmes complexes en physique quantique. A terme, le but est d'aider les chercheurs en physique à construire un ordinateur quantique pour réaliser des calculs combinatoires actuellement hors de portée des ordinateurs classiques.

Intérêt du crowdsourcing/qualités requises : à chaque niveau du jeu, le challenge est de battre le score de l'ordinateur, ce qui signifie réaliser un mouvement plus performant que ceux prévus par les algorithmes de calcul pour amener un électron d'un état à un autre. La seule qualité requise est de l'habileté manuelle. *Quantum Moves*⁵ est un jeu d'adresse.

Challenge/ludification : jeu inaccessible actuellement. ... D'après le site, une nouvelle version est en cours d'élaboration. Dans notre souvenir, la difficulté va croissant, ce qui est stratégiquement bien vu car c'est efficace pour capter l'intérêt du joueur qui passe facilement les premiers niveaux, constitués par un tutoriel, et du coup il a envie de continuer.

4.4. Duolingo

Type de tâche concernée : plutôt microtâche

Remarque : dans l'univers des GWAP, et parmi ceux analysés ici, *Duolingo*⁶ se démarque des autres dans la mesure où il s'agit d'une entreprise à but lucratif : les données fournies par les joueurs ne servent pas à faire progresser la connaissance, à aider des chercheurs, à collaborer à la résolution de problèmes complexes, mais constituent un service facturé à des clients, dans un marché potentiellement très important, celui de la traduction.

Audience/popularité : lancé en juin 2012 sur internet, *Duolingo* est développé pour iPod, iPad et iPhone dès novembre 2012, avant d'être disponible pour les smartphones Android en mai 2013. Le 14 octobre 2013, *Duolingo* fait état de 10 millions d'utilisateurs et annonce alors son partenariat avec les sites d'informations en ligne CNN et BuzzFeed pour la traduction de leurs articles. Ce partenariat est présenté comme un moyen de maintenir la gratuité du site d'apprentissage de langues en ligne. Le fait que Duolingo ait été élue *application de l'année 2013* par Apple fait passer le nombre d'utilisateurs de 16 à 20 millions, et le créateur assure que l'application attire désormais 100 000 utilisateurs de plus chaque jour¹¹

⁵*Quantum Moves* est accessible à l'adresse <http://www.scienceathome.org/index.php/game/play-beta>.

⁶*Duolingo* est accessible à l'adresse <https://www.duolingo.com/>.

Jeux inclassables

But : *Duolingo* est une plate-forme lancée par Luis Von Ahn, le « père » des CAPTCHAS et du concept de GWAP, qui propose l'apprentissage gratuit des langues (anglais uniquement, pour les francophones, et plusieurs langues européennes pour les anglophones). L'apprentissage est ludifié, et le but est double : les données récoltées par *crowdsourcing* sont analysées afin d'améliorer la pédagogie, et les utilisateurs sont invités à tester leurs progrès en faisant de la traduction de *vrais* documents, issus, pour le moment, des sites d'information BuzzFeed et CNN. Certaines sources¹², en évoquant « Duolingo, l'appli qui vous fait vraiment apprendre une langue et vous rend accro, tout en gagnant de l'argent sur votre dos », n'hésitent pas à parler de travail bénévole déguisé, critique déjà relevée ailleurs [Good et Su, 2013] pour la trouvaille initiale de Luis Von Ahn (CAPTCHA et reCAPTCHA). Nous avons testé *Duolingo* pour l'apprentissage de l'anglais.

Intérêt du crowdsourcing pour le problème à résoudre : si on considère que le problème à résoudre est d'éviter de payer des traducteurs professionnels, l'intérêt du *crowdsourcing* tel qu'il est exploité à travers *Duolingo* est évident : en distribuant, à titre d'exercices, les mêmes fragments de texte à traduire à des milliers d'utilisateurs, puis en faisant la synthèse de leurs productions, on est assuré d'obtenir gratuitement une traduction d'un niveau tout à fait correct.

Mais l'application se veut également un outil de réflexion sur la didactique des langues étrangères et utilise le *crowdsourcing* pour tester et comparer en permanence des méthodes d'enseignement, afin d'offrir aux utilisateurs ce qui s'avère le plus efficace. Enfin, pour le fondateur, mettre gratuitement l'apprentissage d'une langue étrangère à la portée de tous est un acte philanthropique, et le bénéfice de la vente de traduction ne sert qu'à maintenir et garantir la gratuité de l'application.

Compétences humaines requises/éducation du joueur : pour produire des données efficaces il semble indispensable que le joueur soit assidu et très motivé par l'apprentissage d'une langue. D'après le concepteur, réaliser la totalité de la progression proposée, en anglais par exemple, équivaut à peu près à avoir le niveau B2 du Cadre Européen Commun de Référence pour les langues (CECR)¹³, soit concrètement être capable de comprendre ce que l'on entend, se débrouiller dans une conversation, pouvoir comprendre un article ou un roman en anglais, ou encore regarder un film en VO.

Interactivité: le joueur/élève est sollicité (voire harcelé...) pour inviter des amis et se mesurer à eux. Sur la page d'accueil, plusieurs boutons incitent à utiliser les réseaux sociaux (Facebook, Google +, Twitter) pour attirer des contacts, chercher des connaissances parmi les gens déjà inscrits, envoyer un mail d'invitation... Une liste de discussions est également accessible et elle semble assez active. Dans le menu figure une rubrique *activité*, qui récapitule toutes les actions du joueur (*a terminé une unité, a traduit x phrases, est maintenant ami avec untel...*). Pour chaque activité recensée, les autres joueurs peuvent *aimer* et/ou *commenter*. L'interactivité entre joueurs est également grandement favorisée au niveau des activités de traduction, puisque dans le module de traduction, on est

Jeux inclassables

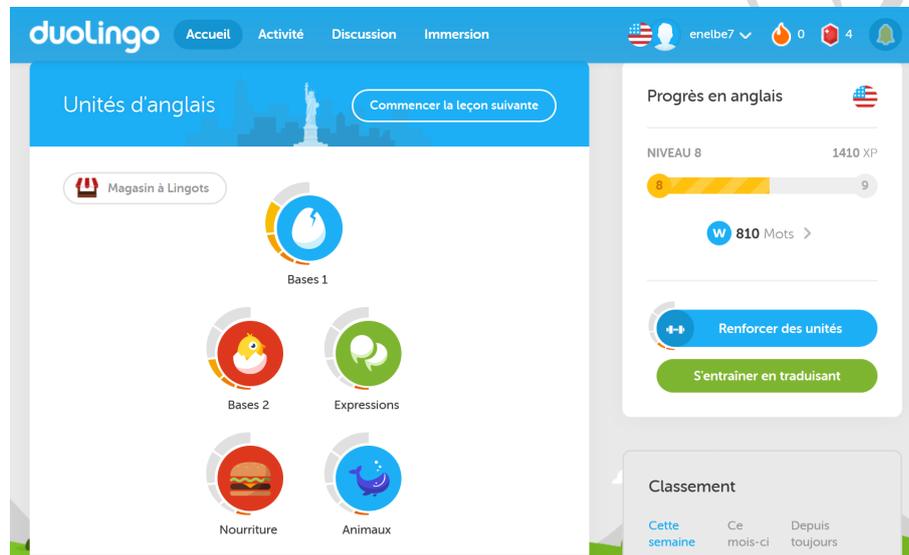


Figure 4.2.* Duolingo. L'écran d'accueil du jeu qui récapitule la progression dans les différents modules d'apprentissage, et propose soit de renforcer ses connaissances, soit de les tester en traduisant des extraits de textes de presse

invité à se prononcer sur les traductions des autres joueurs, et on ne progresse en tant que *traducteur* que grâce aux appréciations positives des autres.

Challenge/Ludification : *Duolingo* mêle dans une interface ludique apprentissage et traduction.

L'apprentissage est organisé en petits *modules* ou unités thématiques, constitués de 1 à 10 *leçons*, chaque *leçon* consistant en une série de 20 questions. L'aspect ludique et motivant est lié à plusieurs éléments :

- la monnaie virtuelle de *Duolingo* est le *lingot* : le joueur gagne des *lingots* en passant au niveau supérieur, en finissant une unité, en finissant une *leçon* sans aucune erreur, en invitant un ami, en étant assidu (c'est-à-dire en jouant au minimum 10 jours consécutifs). Les *lingots* gagnés permettent de faire des achats et transactions diverses dans le *magasin à lingots* ;

- le joueur a droit par défaut à 3 *cœurs* (équivalant à 3 points de vie) au début de chaque *leçon*. A chaque erreur il perd un *cœur*, et un joueur qui perd ses 3 *cœurs* est obligé de recommencer la série d'exercices au début (avec d'autres exercices). Mais il est possible d'acheter un *cœur* de rechange :-) dans le *magasin à lingots* ;

- chaque *leçon* réussie est saluée par un son de fanfare, et permet le passage à la *leçon* suivante ;

Jeux inclassables

- divers bonus peuvent être obtenus en échange de *lingots* dans le *magasin à lingots* : le *gel de série* qui permet un jour d'inactivité sans pénalité, un *quitte ou double* qui permet de doubler le gain en *lingots* lors d'une série de 7 jours ;

- dans le *magasin à lingots*, on trouve également des entraînements, débloquables grâce à des *lingots* : un module d'entraînement chronométré (10 *lingots*), et un quizz pour mesurer ses progrès (25 *lingots*) ;

- enfin, avec 30 *lingots* on débloque des *unités bonus* : *idiomes*, pour apprendre des expressions imagées en anglais, et *drague*, pour apprendre des phrases de drague en anglais. Il n'y a pas de classement proprement dit, mais, à l'accueil, une barre de progression augmente proportionnellement au nombre de *modules* effectués, et un score en XP et un score en mots sont affichés.

La traduction : le joueur est en permanence incité à tester son apprentissage en traduisant des phrases : un bouton *s'entraîner en traduisant* est disponible à tout moment à l'accueil, et à chaque fin de série d'exercices apparaît la mention « tu peux désormais lire x % de vrais articles », suivi de « veux-tu voir ce que ça donne ? » et d'un bouton invitant à traduire un extrait d'article de presse. Le joueur peut également cliquer sur *immersion* et choisir un article à traduire. Il est invité à approuver/désapprouver/modifier les traductions des phrases déjà traduites par d'autres joueurs, ou à initier une nouvelle traduction, qui sera soumise à l'appréciation des autres joueurs. Tout joueur commence au *niveau 1* de traducteur et le passage au niveau suivant est soumis à l'obtention d'au moins 100 votes positifs des autres.

4.5. Le portail ARTigo

But : le nom *ARTigo*⁷ désigne le portail d'un projet destiné à faciliter la navigation et la recherche dans les bases de données relatives aux reproductions d'œuvres d'art, qui peuvent en contenir des millions. Le but des jeux proposés par ce portail est d'associer des mots-clés à des reproductions d'œuvres d'art, afin de faciliter leur recherche. Les reproductions sont celles de la banque de données *Artemis* de l'Institut d'histoire de l'art de l'Université de Munich, qui en contient plus de 25 000.

Intérêt du crowdsourcing pour la tâche à effectuer : l'affectation de mots-clés à chaque reproduction est le seul moyen de rechercher et de retrouver une œuvre, un dessin, un tableau, et vu l'importance des bases de données, le choix de mots-clés pertinents pour cette tâche de reconnaissance est un travail prohibitif en termes de temps et de personnel, et une tâche encore impossible à automatiser, ce qui justifie le recours au *crowdsourcing*. C'est d'autant plus pertinent que l'assignation de mots-clés ne requiert pas une expertise particulière, et que, comme dans beaucoup de jeux d'annotation, ne sont retenus que les mots proposés par au moins deux personnes sans concertation. Les jeux sont directement inspirés du jeu *ESP Game* de Luis Von Ahn [von Ahn et Dabbish, 2004].

⁷Le portail *ARTigo* est accessible à l'adresse <https://www.artigo.org/>.

Jeux inclassables

Compétences humaines requises pour la tâche à effectuer : le choix de mots-clés judicieux ne requiert aucune expertise ni culture en histoire de l'art, mais suppose tout de même un intérêt et une curiosité vis-à-vis de l'art, assortis d'une bonne capacité d'observation.

Le portail *ARTigo* a développé quatre jeux qui ont tous pour but l'association de mots-clés à des œuvres d'art à travers des modalités qui diffèrent légèrement. Pour susciter la motivation, les concepteurs misent clairement sur la combinaison des aspects ludiques (ils se sont efforcés de rendre les jeux excitants) et culturels (on découvre des œuvres d'art), ainsi que sur la valorisation liée au prestige de participer à une description collective d'œuvres d'art.

4.5.1. ARTigo et ARTigo Taboo

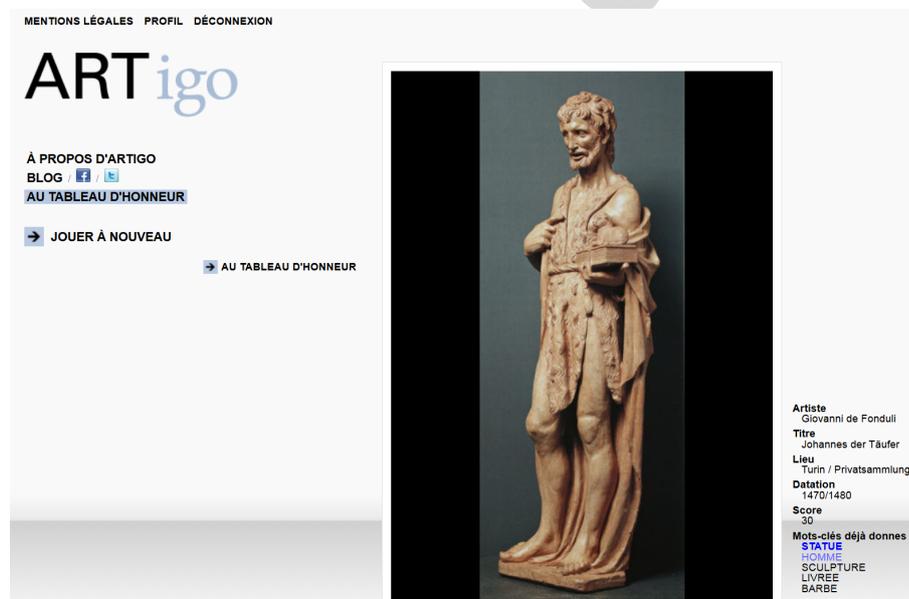


Figure 4.3.* Artigo. A la fin de la partie, on accède pour chaque œuvre d'art au score que l'on a réalisé sur celle-ci (les mots communs sont affichés en bleu), mais aussi aux informations d'identification (titre, auteur, etc.)

Type de tâche concernée : microtâche

Principe : pour *ARTigo*⁸ une œuvre d'art est montrée simultanément à deux personnes, qui doivent, chacune de leur côté, et en temps limité (60 secondes) définir les mots-clés caractérisant le plus précisément possible ce qu'ils voient, et concernant non seulement l'iconographie, mais

⁸*ARTigo* est accessible à l'adresse <https://www.artigo.org/taggingGame.html>.

Jeux inclassables

aussi le style, l'époque et les sentiments qui s'en dégagent. On gagne des points pour chaque mot-clé proposé également par l'autre joueur actif, ou par d'autres joueurs lors de sessions préalables. Dans la version *Taboo*, le principe est le même, assorti de l'interdiction d'utiliser certains mots, déjà attribués [Wieser *et al.*, 2013]. Ces principes ne sont pas sans rappeler ceux rencontrés dans différents jeux du projet JeuxDeMots (développé au chapitre 5).

Ludification : l'aspect jeu est d'abord lié à la surprise : les œuvres d'art à annoter sont tirées au sort et présentées sans aucun lien thématique, ni d'époque ni de nature. Le délai de 60 secondes pour taper les mots-clés qui s'imposent spontanément... ou pas... ajoute du piment à l'exercice, et le fait de voir se colorer en bleu les mots proposés également par l'autre joueur est excitant. Une partie ou *session* comporte 5 œuvres successives et ne dure donc pas plus de 5 minutes. La simplicité et la rapidité font qu'il est possible d'en enchaîner plusieurs, ou en faire une rapidement sans devoir se réapproprier des règles compliquées. Le seul aspect frustrant est de ne pas pouvoir consulter, à l'issue du jeu par exemple, la totalité des mots-clés proposés par l'autre joueur, en particulier quand on n'en a eu aucun en commun. A la fin d'une partie (d'une série de 5 œuvres), on a accès, pour chaque œuvre, à diverses informations (titre, auteur, datation, lieu), ainsi qu'au score que l'on a réalisé sur cette image et au récapitulatif des mots fournis et des mots en commun. En cliquant sur *Au tableau d'honneur* on accède à un score habilement subdivisé en *Meilleurs joueurs du jour/Meilleurs joueurs du mois/Meilleurs joueurs du mois dernier*. Après quelques parties, on a quelques chances, non seulement de figurer dans le premier, voire dans le deuxième, mais également de gagner des places, ce qui est stimulant et invite à rejouer.

Interactivité : l'interactivité est un élément-clé du jeu, puisqu'une partie se déroule en simultané avec un autre partenaire, dont on voit le nombre de propositions augmenter en temps réel, et les propositions identiques s'afficher en bleu vif. On gagne des points en cas de réponses identiques. Les réponses sont également confrontées avec celles données pour cette œuvre lors de sessions précédentes, et en cas de réponse(s) commune(s), elle(s) s'affiche(nt) en bleu pâle et rapporte(nt) également, mais moins de points. Un blog est également accessible, en allemand.

4.5.2. Combino

Type de tâche concernée : microtâche

Principe : une reproduction d'œuvre est proposée, en même temps qu'un ensemble de mots-clés qui s'y rapportent. On dispose de 90 secondes pour associer des mots-clés deux par deux, choisis par *clicks* successifs sur les mots proposés. Les mots proposés sont des mots-clés issus des jeux *ARTigo* et *ARTigo Taboo*, donc des mots validés en tant que mots-clés individuels pertinents car proposés par au moins une paire de joueurs.

Jeux inclassables

Challenge/ludification : on peut obtenir plus de choix en cliquant sur *plus de mots-clés*, et comme pour *Artigo*, on gagne des points dès que l'on propose des combinaisons identiques à celles de l'adversaire (ici le jeu est en version beta, l'adversaire est un programme informatique (bot)), dont on voit en temps réel le nombre de combinaisons proposées augmenter. On gagne également des points, mais moins, si la combinaison a déjà été proposée par un joueur humain lors d'une précédente session. On peut à tout moment passer à une autre œuvre, notamment en cas de manque d'inspiration.

Les auteurs expliquent s'être très largement inspirés du célèbre jeu de Luis Von Ahn, *ESP game* [von Ahn et Dabbish, 2004], que ce soit pour la conception du jeu, ou pour l'étude fine des éléments de ludification propres à doper la motivation du joueur [Störkle, 2012] [Wieser *et al.*, 2013].

4.5.3. Karido

Type de tâche concernée : microtâche

Principe : le but est analogue à celui des jeux précédents, à savoir assigner des mots-clés, avec une nuance cependant : il s'agit de constituer des ensembles de mots-clés associant mots génériques et termes plus spécifiques. Le principe diffère puisque les deux joueurs n'ont pas le même rôle. Neuf reproductions d'œuvres d'art sont présentées simultanément aux deux joueurs, l'un décrit une œuvre au moyen de mots-clés, l'autre doit deviner de laquelle il s'agit parmi les neuf, et ainsi de suite jusqu'à ce que toutes les reproductions aient été traitées. Le jeu permet ainsi de valider en tant que mots-clés ceux qui, isolément, permettent d'identifier une image au sein d'un groupe. Mais le but principal du jeu est d'inciter les joueurs, en leur fournissant des images comportant des points communs, à trouver des mots-clés plus spécifiques afin de permettre une discrimination plus fine entre des images thématiquement ou visuellement proches.

Challenge/Ludification/Interactivité : l'interactivité est vraiment le moteur du jeu puisque tout se passe comme si deux joueurs devaient collaborer efficacement pour gagner des points et être stratégiques dans un rôle comme dans l'autre. La partie commence par le choix du mode de jeu (*limite de temps* ou *limite de coups*) qui peut également influencer le résultat. Dès ce choix effectué, neuf reproductions apparaissent et un rôle est assigné au joueur, soit *descripteur*, soit *devineur*. Le *descripteur* est celui qui choisit une image et tape un mot-clé. Le *devineur* doit alors cliquer sur l'image à laquelle il l'associe. Si c'est la bonne, elle disparaît de l'interface et le *descripteur* en décrit une autre. En cas d'erreur, le *devineur* peut poser une question, et le *descripteur* préciser un autre mot-clé. Quand les neuf reproductions sont reconnues, les rôles sont inversés pour une nouvelle série de neuf images. *Karido*⁹ est de loin le plus amusant des jeux du portail : le challenge qui consiste

⁹*Karido* est accessible à l'adresse <https://www.artigo.org/karido>.

Jeux inclassables

à donner le bon mot-clé pour que le partenaire trouve immédiatement quelle image il désigne est excitant, les limites (de temps ou de nombre de propositions) pimentent la tâche, la nécessité d'être précis et rapide est stimulante, et il faut également être stratégique dans le choix 1) de l'ordre dans lequel on décrit les images (garder les plus difficiles à décrire pour la fin est un gage de réussite) ; 2) du mot-clé, qui doit être le plus précis et pertinent possible dans le cas d'images qui se ressemblent (par exemple, éviter le mot-clé *paysage* si parmi les reproductions il y en a plus d'un).

En fait le joueur humain joue alternativement les rôles de *descripteur* et de *devineur*, mais dans les deux cas joue contre l'ordinateur. Dans le rôle de *descripteur*, le programme réutilise les mots-clés définis par un humain lors d'une précédente session, dans le même ordre et avec les mêmes délais, et dans le rôle de *devineur*, il compare les mots-clés fournis par le joueur avec les mots-clés déjà attribués à l'image dans la base de données et déduit l'image qui convient par calculs de probabilités successifs [Steinmayr *et al.*, 2011]. Signalons que, si les deux joueurs étaient humains, le jeu pourrait être fortement biaisé par la nature des indices donnés en mots-clés (par exemple, le joueur *descripteur* pourrait donner des indications sur la position de l'image sur l'écran, et non sur l'image elle-même).

4.6. Be a martian

Type de tâche concernée : microtâche

But : le programme d'exploration de la planète Mars, mené par la NASA, comprend entre autres trois satellites en orbite autour de Mars et deux robots automobiles, Opportunity et Spirit, qui génèrent une quantité faramineuse de données à dépouiller. La NASA a donc conçu *Be a martian*¹⁰, un espace interactif en ligne à travers lequel le public est invité à participer à trois types de tâches :

- Map Mars : il s'agit d'aligner des photos représentant des fragments de la surface sur un fond de carte plus général afin de réaliser une cartographie précise de la surface de la planète ;

- Count craters : sur des photos de satellites, il faut localiser les cratères en cliquant dessus, ceci afin de fournir des informations sur l'âge relatif de différentes zones de la planète, lequel est proportionnel au nombre de cratères ;

- Tag Mars : Il s'agit de caractériser le plus précisément possible les photos prises par les robots automobiles qui parcourent le sol de Mars, en donnant des informations sur l'aspect du sol (rocheux, sableux...) et du ciel (uniforme, nuageux, étoilé) pour classer la base de données d'images de Mars.

¹⁰*Be a martian* est accessible à l'adresse <http://beamartian.jpl.nasa.gov/welcome>.

Jeux inclassables

Intérêt du crowdsourcing : l'intérêt du *crowdsourcing* est lié à la fois à l'énormité du volume de données à traiter, et à la difficulté d'automatiser la caractérisation fine des éléments visibles sur les images. La tâche demandée est typiquement de type microtâche, et ne nécessite aucun savoir-faire particulier.

Challenge/ludification/interactivité : *Be a martian* est typiquement une activité de *crowdsourcing* qui tente de se faire passer pour un jeu en multipliant les emprunts à l'univers ludique. Le décor imite le centre de pilotage d'un vaisseau spatial, où les activités sont proposées *via* des écrans de contrôle. A l'inscription, on choisit un *avatar* et un *pseudo*, et les différentes contributions réalisées dans les trois domaines donnent lieu à des titres, distinctions, badges, points et prix divers...

Pas réellement d'interactivité, on joue seul, mais beaucoup d'informations scientifiques sur la planète Mars et sur le programme de la NASA sont dispensées à travers le *Citizenship Hall*, qui héberge les tâches de *crowdsourcing* et d'autres activités de vulgarisation plus ou moins ludifiées (quizz, tourisme sur la planète Mars, envoi de cartes postales... au robot automobile d'exploration Spirit!...)



Figure 4.4.* *Be a martian*. L'écran d'accueil propose diverses activités de crowdsourcing relatives au dépouillement de la masse de données recueillies par les satellites et les robots d'exploration

4.7. Akinator, le génie du Web

Type de tâche concernée : microtâche

Lancement : 2007

Ce jeu a été initialement conçu par Arnaud Megret, en amateur, en 2007, mais *Akinator*¹¹ n'est réellement devenu célèbre qu'en 2009 avec son lancement sur smartphones. La figure 4.5 reproduit l'écran d'accueil.

But : c'est avant tout un jeu, d'où son développement sur smartphones. Il a pour but de collecter les caractéristiques déterminantes/discriminantes de personnages, et depuis une extension récente, de n'importe quel type d'entité.



Figure 4.5.* Akinator : Partie principale de l'écran d'accueil. Cet écran montre une interface particulièrement ludique : le jeu est disponible en plusieurs langues (14 actuellement), il est possible d'y jouer sur différents types de smartphones et il existe une version spécifique pour les enfants. Les résultats des 10 dernières parties sont également affichés, montrant ainsi le taux de réussite d'Akinator (90 % sur l'écran présenté ici), sur des personnages parfois peu connus ou peu médiatisés.

Une partie se déroule entre le génie *Akinator* et un joueur. Il s'agit d'un jeu de devinettes où le joueur pense à un personnage, réel ou fictif, et *Akinator* va tenter de le deviner en lui posant des questions. Le jeu possède une base de données qui regroupe environ 100 000 personnages avec

¹¹*Akinator* est accessible à l'adresse <http://fr.akinator.com>.

Jeux inclassables

leurs caractéristiques ; il choisit les questions de façon à éliminer, à chaque réponse du joueur, le plus de possibilités (par exemple : « le personnage est-il vivant ? », « est-ce un homme ? », « est-il européen ? »). Les réponses du joueur sont nécessairement : *Oui, Non, Ne sais pas, Probablement, Probablement pas*. Les questions sont organisées et posées de manière logique, comme le ferait un humain (par exemple : après une réponse négative à la question « est-il européen ? », *Akinator* pourra demander si le personnage est américain). Au bout d'une petite vingtaine de questions, la base d'*Akinator* ne contient généralement plus qu'un seul personnage correspondant aux réponses du joueur ; ce dernier peut alors faire une proposition. Si celle-ci est incorrecte, il propose au joueur de continuer en lui posant une nouvelle série de questions (qui ont souvent pour but d'élargir un peu la recherche, en supposant que le joueur a pu se tromper quelques fois dans ses réponses). Si après ces questions supplémentaires *Akinator* se trompe, ou ne trouve pas le personnage, il demande au joueur de lui indiquer quel était le personnage, et ce qui le caractérise : il enrichit ainsi sa base de connaissances. Le jeu repose donc sur l'honnêteté des joueurs, mais reconnaissons qu'il est peu motivant de faire deviner un personnage avec des indices volontairement erronés ! Quel intérêt y aurait-il à fournir de tels indices pour *perdre Akinator* ? (Le taux de *trollage* doit effectivement être très faible.) Le joueur répond majoritairement par Oui ou Non, mais il a la possibilité d'avouer son ignorance, ou sa non-certitude, concernant la question posée sur le personnage auquel il pense. La figure 4.6 reproduit quelques écrans d'une partie.

Akinator affiche sur son écran d'accueil (figure 4.5) les résultats des dix dernières parties : généralement son taux de réussite est de 9 ou 10/10. A ce jour, plus de 200 millions de parties ont été jouées et, quotidiennement, plusieurs dizaines de milliers de parties sont jouées en France, majoritairement sur smartphone. Reconnaissons qu'avant tout qu'il s'agit d'un jeu ludique, la partie GWAP étant réduite aux seuls cas où *Akinator* ne connaissait pas le personnage auquel pensait le joueur.

Une nouvelle version d'*Akinator* propose de deviner des objets, animaux ou concepts, mais la plupart des joueurs préfèrent lui faire deviner des personnages. Il existe également une version *Akinator Kids* destinée aux enfants, disponible en mode GWAP, tout comme la version classique, mais également en mode jeu éducatif où c'est *Akinator* qui fournit des indices pour faire deviner des personnages célèbres au jeune joueur.

On peut toutefois légitimement se demander si *Akinator* est effectivement un GWAP ou un simple jeu de devinettes. Les données collectées ne sont pas librement disponibles et leur forme exacte ainsi que ce qu'il serait possible d'en faire restent inconnus du grand public. Sur le site Web d'*Akinator*, le jeu est davantage une vitrine pour les techniques et algorithmes qu'il utilise que pour son intérêt dans une éventuelle collecte de données.

Jeux inclassables

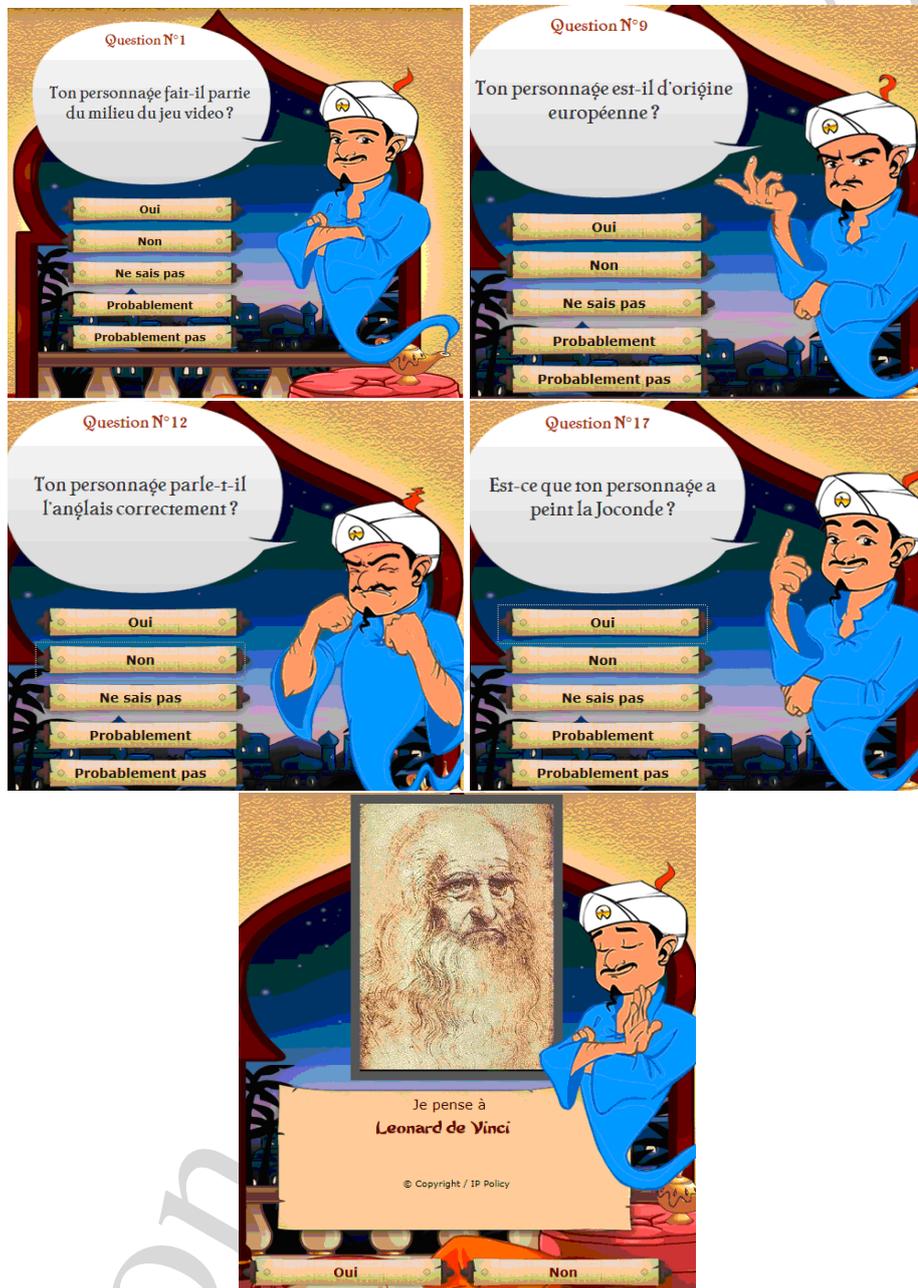


Figure 4.6.* Ecrans d'Akinator. Succession de quelques images écrans montrant des exemples de questions posées par le génie Akinator. Observez ses différentes mimiques, en particulier sa satisfaction lorsqu'il pense avoir trouvé

Jeux inclassables

Notes

¹¹<http://fr.wikipedia.org/wiki/Duolingo>.

¹²<http://www.slate.fr/story/83419/duolingo-appli-langue-rend-accro>.

¹³http://fr.wikipedia.org/wiki/Cadre_europeen_commun_de_reference_pour_les_langues.

Jeux inclassables

Version auteurs, dec. 2014

Le projet *JeuxDeMots* – des jeux et des mots

Le projet *JeuxDeMots*¹, décrit à l'origine dans [Lafourcade, 2007] puis beaucoup plus en détail dans [Lafourcade et Joubert, 2010], a pour but la constitution d'un réseau lexico-sémantique de grande taille pour le français. Il s'agit d'une collection de jeux et contre-jeux qui concourent ensemble, d'une manière ou d'une autre à cet objectif.

La présentation et l'analyse des jeux du projet *JeuxDeMots* permettent de mettre en évidence les qualités que doit avoir un GWAP, en particulier les conditions qu'il doit remplir pour être efficace. De plus, le fait que le projet ait été initié en 2007 donne le recul nécessaire pour en évaluer l'impact effectif en termes de quantité et de qualité des ressources collectées. Nous évoquons les limites et les biais de chacun des jeux du projet et montrons comment les contre-jeux permettent de valider les données et de compenser au moins partiellement les biais induits par le jeu principal.

5.1. L'objet à construire, un réseau lexical

Les réseaux lexicaux, qu'ils soient généraux ou spécialisés, constituent, en particulier depuis l'émergence de WordNet [Miller *et al.*, 1990] (mais également de EuroWordNet [Vossen, 1998], entre autres) une ressource particulièrement précieuse pour les applications de l'informatique lin-

¹*JeuxDeMots* est accessible à <http://www.jeuxdemots.org>.

guistique, dont notamment la désambiguïsation lexicale [Véronis et Ide, 1990]. Dans la plupart des réseaux lexicaux disponibles, les relations ne sont pas pondérées, c'est-à-dire qu'elles sont simplement énumérées, sans aucune indication quant à leur *force* ou leur fréquence. La pondération permet de faire la distinction entre les relations fortes et celles qui sont plus anecdotiques, ces dernières n'étant pas nécessairement moins importantes dans les processus de TALN. Les algorithmes de propagation en désambiguïsation peuvent tirer profit de telles pondérations. Cependant, la question des facteurs qui déterminent le poids d'une relation reste délicate : il est fortement lié à la fréquence d'usage de la relation, mais pas uniquement. Ce type de ressource peut être construit avec plus ou moins de bonheur manuellement et/ou par croisement de données avec des procédures plus ou moins automatisées suivi d'une vérification, manuelle également [Sagot et Fiser, 2008].

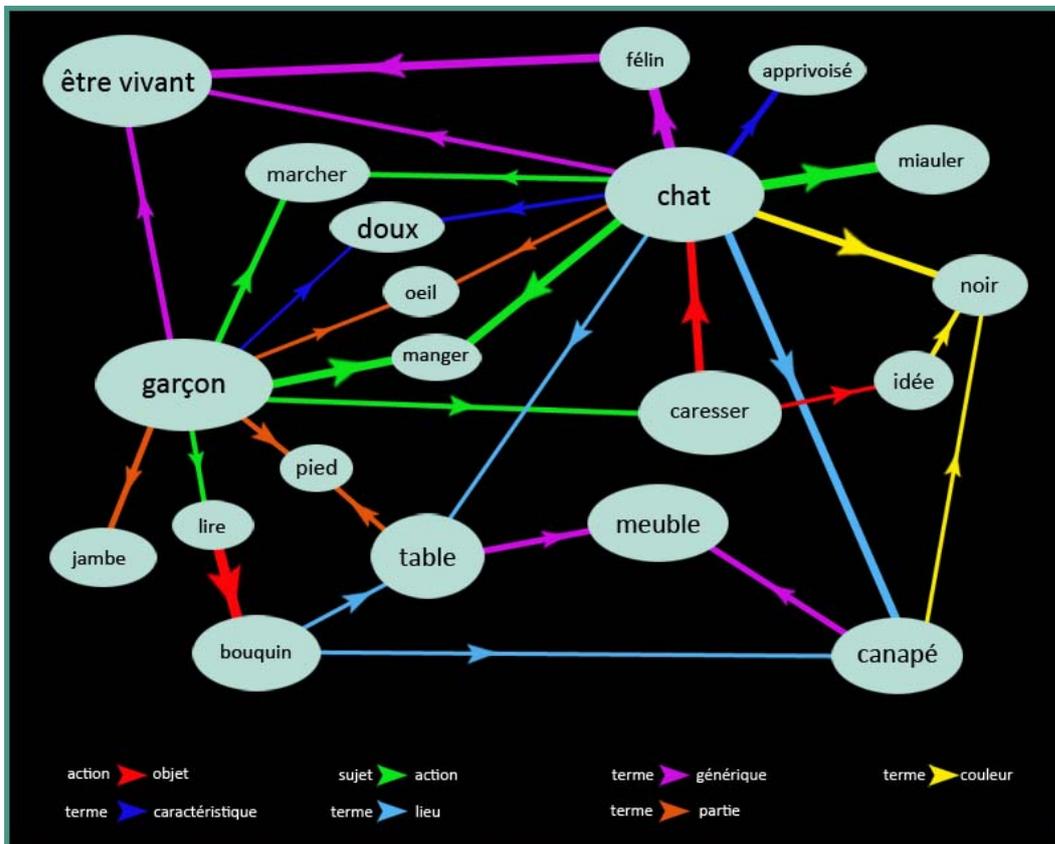


Figure 5.1.* Illustration d'un réseau lexical. Schéma d'une portion de réseau lexical mettant en évidence quelques relations

Le projet *JeuxDeMots*

De plus, les réseaux lexico-sémantiques contiennent souvent un nombre variable de types de relations, généralement ontologiques (hyperonymes) ou lexicales (synonymes, antonymes). HowNet [Dong et Dong, 2006] est un exemple d'une telle ressource lexicale (chinois et anglais) construite manuellement et présentant de nombreux types de relations. Cyc [Lenat, 1995] concrétise un effort majeur de construction manuelle d'une base de connaissances générales fondées sur la logique, dont la forme même peut être assimilable à un réseau sémantique, mais pas ou peu lexical. Enfin, certains réseaux contiennent également des relations associatives monolingues ou multilingues. Par exemple, BabelNet est un réseau multilingue de grande taille [Navigli et Ponzetto, 2010] issu d'une extraction/compilation automatique des informations de Wikipédia.

5.2. *JeuxDeMots*, un jeu associatif

JDM, le tout premier jeu du projet, a été mis en service en juillet 2007, dans le but de constituer un réseau lexical pour le français. Ce réseau a été amorcé avec environ 150 000 termes sans aucune relation entre eux. Au bout de sept ans et 1 426 603 parties jouées, au 1^{er} octobre 2014, le réseau contient 420 000 termes et plus de 12 millions de relations.

DONNER DES ASSOCIATIONS D'IDEES AVEC LE TERME QUI SUIT :

... record à battre de 450 Cr.

meuble

Temps
7 s

mettre un terme ici

OK

Dernier terme proposé : meuble en kit • supprimer

Ce terme a plusieurs sens ou il en manque ? [Demandez de l'aide à vos amis](#)

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 demander de l'aide à vos amis. Vous pouvez supprimer un mot proposé en cliquant dessus dans la liste affichée à droite.

meuble en kit
Ikea
ameublement
< commode (meuble)
< fauteuil (siège)
fauteuil
< lit (meuble)
lit
> tabouret >
< mobilier (meuble)
< bois (matière)
bois >
< en bois (matériau)
en bois >
appartement
armoire >
< table (meuble)
table >
mobilier >
terre >

0:17 s

meuble en kit • Ikea • ameublement • commode (meuble) • fauteuil (siège) • fauteuil • lit (meuble) • lit • tabouret • mobilier (meuble) • bois (matière) • bois • en bois (matériau) • en bois • appartement • armoire • table (meuble) • table • mobilier • terre

meuble en kit • Ikea • ameublement • commode (meuble) • fauteuil (siège) • fauteuil • lit (meuble) • lit • tabouret • mobilier (meuble) • bois (matière) • bois • en bois (matériau) • en bois • appartement • armoire • table (meuble) • table • mobilier • terre

meuble en kit • Ikea • ameublement • commode (meuble) • fauteuil (siège) • fauteuil • lit (meuble) • lit • tabouret • mobilier (meuble) • bois (matière) • bois • en bois (matériau) • en bois • appartement • armoire • table (meuble) • table • mobilier • terre

Figure 5.2.* Une partie typique de *JeuxDeMots*. Le joueur doit proposer des termes qu'il associe au mot cible *meuble*. Ses propositions apparaissent à droite (en vert) au fur et à mesure qu'il les saisit

Le projet *JeuxDeMots*

JDM est un jeu dit *ouvert*, les joueurs devant entrer en un temps limité des propositions à partir d'un terme cible et d'une consigne. Cette modalité d'interaction réduit fortement les scénarii d'utilisation, mais est de loin la plus riche de l'ensemble des jeux proposés. Une partie de JDM, sur un terme et un type de relation, se joue anonymement et de façon asynchrone entre deux joueurs. Pour des raisons de sécurité et de qualité de la ressource obtenue, nous ne mémorisons que les réponses communes aux deux joueurs. Les relations ainsi acquises sont pondérées en fonction du nombre de paires de joueurs qui les ont proposées.

La figure 5.2 présente une partie typique de *JeuxDeMots* où il est demandé d'énumérer ce que peut évoquer le terme *meuble*. Le joueur a fait un bon nombre de propositions et, pour certaines d'entre elles, a sélectionné l'usage qui lui semblait être pertinent. Par exemple, parmi les usages possibles de *commode*, il a choisi *commode (meuble)*. Le résultat de la partie est présenté dans la figure 5.3. La liste des termes proposés en commun avec l'autre joueur est affichée. Chacun des deux joueurs (l'actif comme le passif) récolte les gains annoncés. Le score étant élevé, un certain nombre d'actions sont possibles et constituent une récompense supplémentaire pour le joueur actif.



The screenshot shows the game interface for the word "meuble". At the top right, there are two scores: 102 093 and 317. The word "meuble" is displayed in large pink letters. Below it are two lists of suggestions:

Réponses données par kaput : meuble en kit • Ikea • ameublement • commode (meuble) • fauteuil (siège) • fauteuil • lit • lit (meuble) • tabouret • mobilier (meuble) • bois (matière) • bois • en bois (matériau) • en bois • appartement • armoire • table (meuble) • table • mobilier • terre

Réponses données par Eolidou : coin • vaisselier • meuble de balcon • vernir • scier • guéridon • vieux • pièce de mobilier • table à manger • customisé • berceau • meuble de laque • meuble en chêne massif • table basse • zinc (comptoir) • secrétaire • bois • confortable • bois (matière) • wacapou • décaper • table à tréteaux • zinc (comptoir) • étagère • objet manufacturé • meuble de salon • bien meuble • servante (petite table) • meuble en chêne • meuble moderne • meuble caisse • meuble en noyer • meuble en chêne massif • meuble ancien • terre meuble • vaisselier • meuble de cuisine • meuble de cuisine • canapé • meuble de salle à manger • bar • buffet • fauteuil • placard • bureau • commode • lit • mobilier • bois • chaise • table • armoire

Below the lists, there is a line of common terms: **fauteuil • lit • bois (matière)** (with a star icon) • bois • armoire • table • mobilier

At the bottom, a blue banner reads: **Vous gagnez 275 crédits et 10 points d'honneur**. Below this is a Facebook "J'aime" button with the text "Soyez le premier de vos amis à indiquer que vous aimez ça."

Figure 5.3.* Le résultat de la partie précédente. Le joueur actif Kaput a sept termes en commun avec le joueur passif Eolidou. Les deux joueurs gagnent 275 crédits et 10 points d'honneur

Une partie considérée comme ratée peut faire l'objet d'un procès. Intenté par le joueur actif, un procès consiste à rendre publiques les réponses du joueur passif et à exposer les griefs du plaignant.

Le projet *JeuxDeMots*

Les autres joueurs inscrits, prévenus par un post-it en page d'accueil du jeu, peuvent alors s'exprimer en votant pour ou contre la partie mise en accusation, ou encore en s'abstenant. Au bout d'une semaine, le procès est clos et le gagnant (l'accusateur ou l'accusé) reçoit un dédommagement symbolique. Grâce à l'espace de discussion, où chacun peut argumenter dans un sens ou dans l'autre et justifier son vote, les procès ont un effet pédagogique notoire, en particulier concernant les relations difficiles.

Les joueurs ont également la possibilité d'offrir des cadeaux aux autres participants. Un cadeau consiste en la possibilité de jouer une partie sur un couple terme/relation choisi par celui qui le fait. La motivation est ici essentiellement d'ordre social : un joueur qui offre des parties intéressantes est bien perçu par la communauté. Le développement du réseau lexical s'en trouve accéléré et bonifié, l'immense majorité des joueurs s'abstenant d'offrir des couples terme/relation sans intérêt. L'analyse du comportement des joueurs offrant des cadeaux montre que la pression sociale (*via* le forum ou les procès) empêche les comportements stériles ou d'anti-jeu.

Outre les parties standard, *JeuxDeMots* offre une multitude de variantes de jeu, qui permettent de relancer l'intérêt du joueur afin de rompre la monotonie et d'éviter qu'il ne se lasse. Entre autres, il est possible de :

- jouer sur les thématiques de son choix, ou sélectionner celles des autres joueurs ;
- jouer sur des termes *faciles* ou *difficiles*, le critère de difficulté étant automatiquement déterminé par l'état du réseau lexical ;
- lancer des *loteries*, qui permettent d'apparier deux parties au hasard (une partie du joueur connecté avec celle d'un autre joueur) ; en fonction du nombre de termes en commun (*intersections*) on gagne des lots de diverses natures ;
- tenter de voler des mots à d'autres joueurs, les provoquer en duel, effectuer des missions, etc.

L'ensemble de ces modes de jeux a pour but d'augmenter l'activité de renseignement du réseau. De nombreuses actions sont *payantes* (dans les monnaies virtuelles du jeu, les *crédits* et les *pilules bleues*) et visent implicitement, en incitant le joueur à dépenser son butin, à créer une situation de *famine*, dont le joueur ne pourra sortir qu'en faisant des parties standard pour refaire son stock de *crédits* et *pilules bleues*.

Les parties en elles-mêmes sont strictement coopératives car les deux joueurs (actif et passif) ont exactement les mêmes gains. Cependant, de nombreux éléments du jeu incitent à la compétition, en particulier la présence de multiples classements, la diversité des critères de classement incitant les joueurs à jouer, à la fois le mieux, et le plus souvent possible. Parmi ceux-ci, le nombre de mots possédés (puisque l'ambition affichée du jeu est de *collectionner* des mots) peut s'avérer particulièrement motivant. La concurrence acharnée entre joueurs est encouragée dans la mesure où, impliquant de jouer mieux que l'autre, elle est extrêmement enrichissante pour le réseau. Parallèle-

Le projet *JeuxDeMots*

ment, la stratégie consistant à mal jouer pour faire chuter l'autre n'est pas viable puisqu'elle conduit à sacrifier ses propres parties et donc à chuter également dans les classements. Ainsi, mal jouer ne peut en aucun cas permettre d'être mieux classé ou d'obtenir un avantage quelconque.

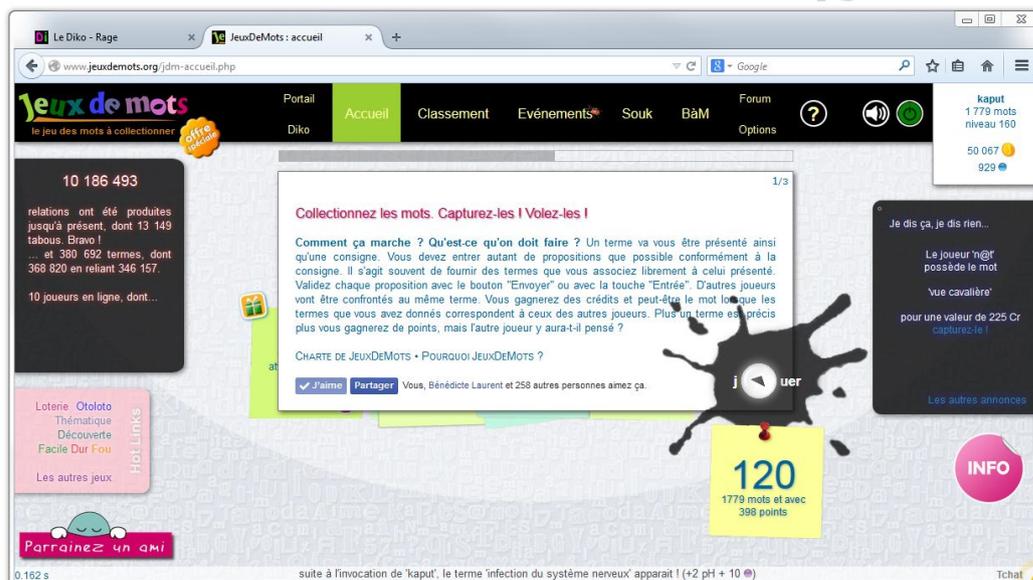


Figure 5.4.* Page d'accueil du jeu *JeuxDeMots*

On voit que *JeuxDeMots* est clairement un jeu de coopération/compétition. En jouant bien, non seulement on augmente son propre score global, mais on influence favorablement ceux des autres joueurs. La diversité des critères de classement (points gagnés, nombre de parties jouées, efficacité, etc.) renforce l'activité tout en incitant chaque joueur à jouer le mieux possible en privilégiant telle ou telle stratégie ou mode de jeu en fonction du type de classement qui le motive le plus.

Les autres (contre-)jeux sont dits *fermés* dans la mesure où une ou plusieurs propositions sont affichées et doivent être choisies ou validées. De façon générale, les parties sont plus courtes que celles de JDM et sont jouables *via* des smartphones ou des tablettes tactiles. Ce dernier point est loin d'être anecdotique dans la mesure où bon nombre de joueurs ont indiqué qu'ils jouaient à ces jeux dans des situations d'attente (transports en commun, etc.). Le jeu *Totaki* reste toutefois un jeu partiellement ouvert, car si le système ne trouve pas la bonne réponse, le joueur est invité à faire une proposition.

5.3. *PtiClic*, un jeu d'attribution

Le jeu *PtiClic* [Lafourcade et Zampa, 2009], lancé un an après JDM, a pour but la consolidation du réseau créé grâce à JDM : il permet de renforcer ou préciser les relations existantes. *PtiClic* est un jeu fermé d'attributions par glisser-déposer : un mot cible et un nuage de termes liés s'affichent, et le joueur doit déplacer un par un les termes du nuage dans les catégories qui lui semblent pertinentes relativement au mot cible. Par exemple (figure 5.5), pour le mot cible *restaurant*, le terme *commerce* doit-il plutôt être considéré comme un hyperonyme ou un synonyme ? Lors d'une partie, certains termes du nuage ne sont pas attribuables à une relation et doivent être ignorés. Le nuage de termes est créé en partie sur la base du réseau lexical JDM mais également *via* une analyse LSA d'un gros corpus (journalistique et textes d'enfants).

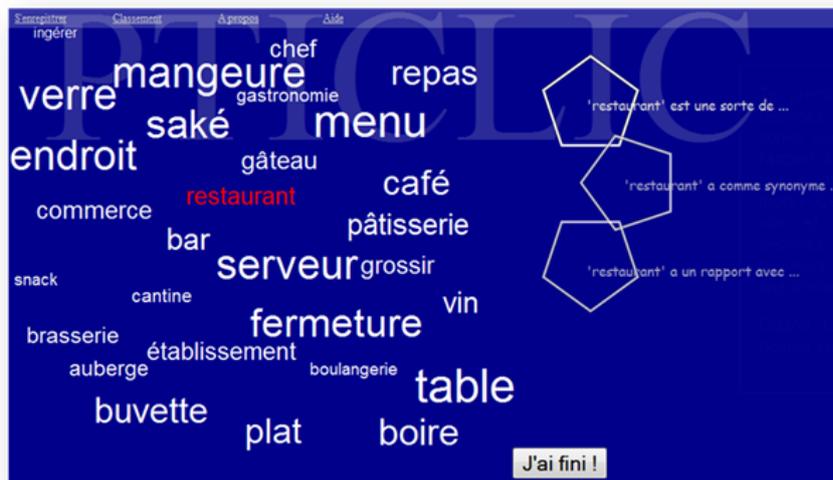


Figure 5.5.* Partie typique de *PtiClic*. Pour le mot cible *restaurant*, les trois relations cibles proposées sont l'hyperonymie, la synonymie et l'association d'idées

Une partie de *PtiClic* n'est pas limitée dans sa durée, mais en pratique elle dure moins longtemps (entre 30 et 45 secondes) qu'une partie de JDM, mais plus qu'une partie de jeu à clics. La modalité d'interaction rend possible l'utilisation de tablettes ou de smartphones. Au premier janvier 2013, environ 60 000 parties de *PtiClic* ont été jouées depuis son lancement.

L'idée de *PtiClic* est de proposer, pour le mot cible, une série de termes qui lui sont associés dans le réseau lexical JDM par une relation très générale (*idées associées*) et de demander au joueur de définir pour ces termes des relations plus précises avec le mot cible, comme l'hyperonymie (générique), l'hyponymie (spécifique), l'holonymie (tout), la méronymie (partie), etc. Ce sont donc des

Le projet JeuxDeMots

relations ontologiques qui sont proposées comme catégories, les relations strictement lexicales (par exemple, *termes de la même famille*) se prêtant beaucoup moins bien à ce type d'exercice.

La relation *idées associées* reste proposée comme choix possible lors des parties. En effet, certains termes du nuage sont extraits *via* des méthodes basées sur LSA ou dans un voisinage plus ou moins éloigné du terme cible dans le réseau. Ce faisant, il n'est pas certain que ces termes-là aient un rapport avec le terme cible.

Une extension de *PtiClic* [Zampa et Lafourcade, 2010] a permis de collecter des informations sur l'âge d'acquisition du vocabulaire chez les jeunes locuteurs. Ce type d'approche peut et doit se faire dans un contexte pédagogique et trouve son utilité en analyse de la difficulté lexicale de textes.

5.4. Totaki, un jeu de devinettes

Totaki, lancé début 2010, peut être considéré comme un jeu de devinettes, mais également comme un outil d'aide au problème du Mot sur le Bout de la Langue (MBL) [Joubert *et al.*, 2011]. Initialement, *Totaki* a été pensé et conçu comme une modalité d'évaluation qualitative du réseau lexical JDM. Cependant, le succès de ce jeu et le fait que beaucoup de joueurs cherchent à *piéger Totaki*, ont également conduit à des modalités de jeu qui enrichissent le réseau.

Le principe du jeu consiste à faire deviner un mot au système en lui proposant des indices. Après chaque indice, le système propose une réponse possible. Les indices peuvent être typés, sinon par défaut on les considère comme simplement associés au mot à trouver. Le type d'un indice correspond à un des types de relations existant dans le réseau (par exemple, on peut donner comme un indice la relation générique *:isa animal*, au lieu de donner *animal* tout court).

| Vos indices | Mes réponses |
|-----------------|-------------------------|
| pommes de terre | pommes de terre à l'eau |
| pommes vapeur | pommes |
| chair ferme | Belle de Fontenay |

Figure 5.6.* Partie typique de Totaki. Pour faire deviner Belle de Fontenay à Totaki, trois indices ont suffi...

Le projet JeuxDeMots

Données Totaki (24544 données de taille 245)

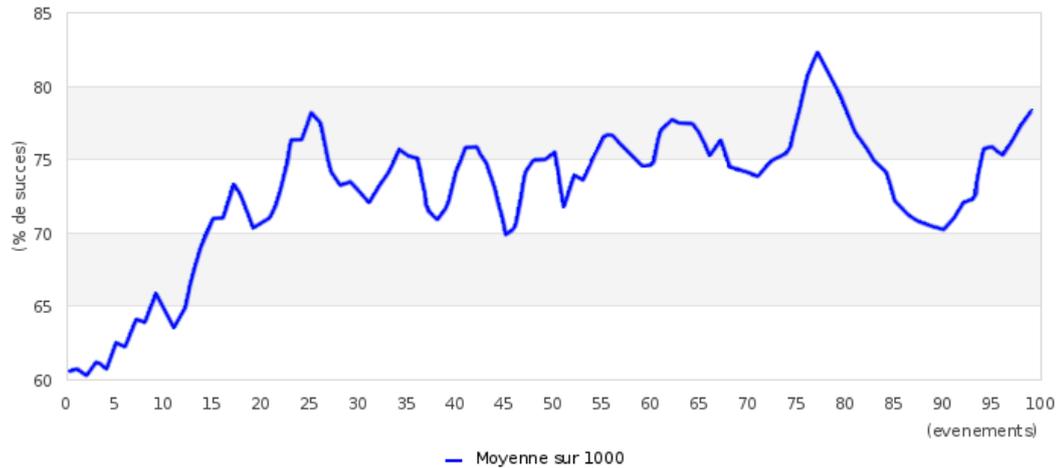


Figure 5.7.* Taux de réussite de Totaki. Evolution du taux de réussite de Totaki à deviner correctement le terme ciblé par le joueur

L'évaluation du réseau lexico-sémantique *via Totaki* a été présentée de façon très détaillée dans [Joubert et Lafourcade, 2012]. Le résultat le plus notable est que *Totaki* devine le terme ciblé par le joueur dans environ 3 cas sur 4. Sous les mêmes conditions, les joueurs humains y parviennent dans 48% des cas. Outre une évaluation des données obtenues *via* les autres jeux, *Totaki* permet aussi un renforcement du réseau et des détections d'incohérences par un mécanisme d'apprentissage. Depuis son lancement, plus de 34 000 parties de *Totaki* ont été jouées (l'enregistrement à des fins statistiques n'ayant été réalisé qu'à partir de la 10 000^e).

5.5. Quelques jeux à votes

Le principe d'un jeu à votes est de proposer à l'utilisateur de se prononcer sur un énoncé relatif aux données lexicales. Il s'agit d'une forme de sondage dont la présentation prend un tour ludique. Le vote se matérialise en général par un clic sur une proposition parmi plusieurs (on parlera de *jeu à clics* si on s'intéresse à la modalité). Ce type de jeu a l'avantage d'être très rapide pour le joueur et de fournir une granularité très fine en termes de modification de la ressource lexicale. Suite à son choix, le joueur est informé de la manière dont ce dernier le situe par rapport aux autres joueurs, autrement dit s'il a répondu comme la majorité des gens ou, au contraire, si le choix qu'il

Le projet JeuxDeMots

a fait est statistiquement minoritaire. Cette *réponse* du jeu est la forme de récompense qu'attend le joueur. L'objectif implicite du jeu, implicite car non annoncé explicitement aux joueurs, est d'obtenir des réponses honnêtes aux questions, afin de faire idéalement émerger un consensus (si celui-ci est possible). Les jeux à votes sont en général de bons contre-jeux aux jeux ouverts dans la mesure où d'une part, ils ne peuvent donner lieu à des *réponses forcées* (on parle de réponse forcée quand le joueur peut répondre n'importe quoi parce qu'il dispose d'un champ de texte libre et qu'il peut être tenté de répondre pour répondre) d'autre part, les joueurs sont relativement peu indulgents avec les propositions limites et/ou incohérentes : ils vont quasi systématiquement dépister *via* les jeux à votes les réponses forcées issues des jeux ouverts.

AskIt² : Le jeu AskIt est un *jeu à clics* (sans utilisation du clavier) portant sur les liens entre des usages de termes et d'autres termes : par une série de questions dont les réponses possibles sont simplement *oui*, *non* et *possible mais bof*, il permet un enrichissement et une consolidation du réseau, en particulier en ce qui concerne les raffinements des termes polysémiques.

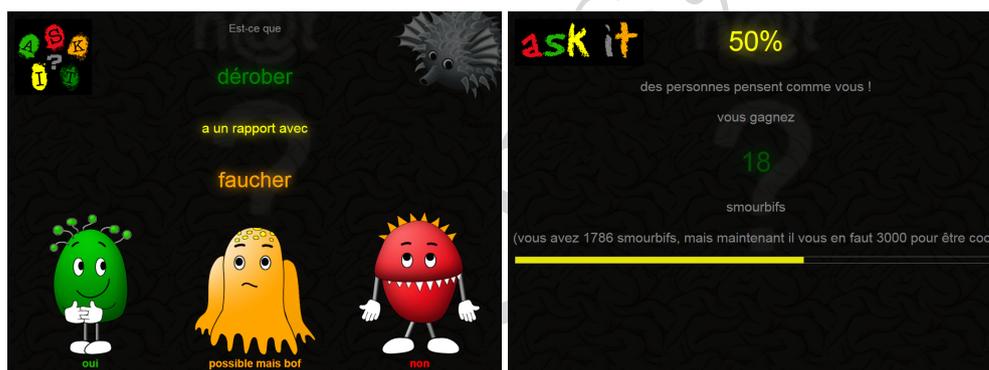


Figure 5.8.* Partie typique d'AskIt. A gauche l'énoncé sur lequel le joueur doit se prononcer en répondant oui, non, ou possible mais bof, à droite le résultat : affichage du pourcentage de joueurs du même avis, et de la barre de progression du score associé

AskIt est un jeu de consensus fermé et rapide. L'analyse des motivations des joueurs montre que l'excitation/frustration suite à une partie provient de l'espérance d'avoir répondu – ou non – comme la majorité. Le gain symbolique en points est matérialisé par une barre de progression. L'objectif implicite de la remplir est une incitation psychologique forte à jouer honnêtement. Le taux de trollage (joueurs répondant volontairement quelque chose d'erroné) a été estimé à moins de 1% de l'ensemble de plus de 500 000 parties jouées à ce jour. L'algorithme de sélection des questions est double :

²AskIt est accessible à l'adresse <http://www.jeuxdemots.org/askit.php>.

Le projet *JeuxDeMots*

- sélection par triangulation dans le réseau :

Il s'agit ici de propager des relations par induction avec des schémas de type :

$$A \ r_1 \ B \quad \& \quad B \ r_2 \ C \quad \Rightarrow? \quad A \ r_3 \ C$$

avec des sélections particulières de triplet (r_1 , r_2 , r_3). Par exemple, si *canari isa oiseau* (un canari est un oiseau) et *oiseau agent chanter* (un oiseau peut chanter) alors avons-nous *canari agent chanter* (un canari peut chanter) ? Nous ferons remarquer que les réponses à ce genre de questions sont loin d'être systématiquement positives. En effet, un contre-exemple serait : *cigogne isa oiseau* et *oiseau agent chanter* \Rightarrow * *cigogne agent chanter*³ (les cigognes font partie des très rares oiseaux qui ne chantent pas). De plus, nous nous focalisons particulièrement sur la relation de raffinement, rendant ainsi possible le basculement de relations d'un terme polysémique vers un ou plusieurs de ses usages :

frégate \rightarrow *raffinement* \rightarrow *frégate* (oiseau)
& *frégate* \rightarrow *partof* \rightarrow *coque* (*frégate* a pour partie *coque*)
 \Rightarrow * *frégate* (oiseau)³ \rightarrow *partof* \rightarrow *coque* (Il est FAUX de dire que *frégate* (oiseau) a comme partie *coque*)
frégate \rightarrow *raffinement* \rightarrow *frégate* (navire)
& *frégate* \rightarrow *partof* \rightarrow *coque* (*frégate* a pour partie *coque*)
 \Rightarrow *frégate* (navire) \rightarrow *partof* \rightarrow *coque* (*frégate* (navire) a bien comme partie *coque*)

- sélection de relations faibles :

La seconde méthode consiste à demander confirmation (ou infirmation) pour des occurrences de relations ayant un poids faible, et donc sujettes à caution.

La triangulation est une approche simpliste qui donne environ 60 % de résultats corrects. Ce qui veut dire qu'environ 40 % des propositions faites aux joueurs sont erronées. Nous ferons remarquer que cela est souhaitable et volontaire. En effet, un algorithme *trop bon*, qui générerait des propositions donnant lieu à une majorité de réponses positives, serait ennuyeux pour le joueur. La proportion importante de propositions fausses mais intéressantes entretient l'attention du joueur et son intérêt pour le jeu.

Nous essayons par ailleurs de piéger le joueur en inversant des propositions avérées correctes. Cela permet d'amener un peu plus de variété dans les réponses et surtout de détecter les joueurs qui répondent au hasard ou systématiquement.

Une réponse *non* diminue légèrement le poids de la relation en jeu. Le poids de certaines relations peut donc devenir négatif (elles sont alors dites *inhibitrices*). Par exemple, *autruche agent voler* (une

³* désigne une assertion fausse.

Le projet *JeuxDeMots*

autruche peut voler) a un poids négatif de -45 dans la base (au moment de l'écriture de ce chapitre). Ces relations inhibitrices sont particulièrement intéressantes en analyse sémantique, car en jouant un rôle d'interdiction elle arrêtent le signal de propagation sur un réseau d'occurrences de termes. Environ 10 000 relations inhibitrices ont été obtenues grâce à *AskIt* en 12 mois.

AskIt possède un certain nombre de caractéristiques notables, à savoir :

- c'est un jeu addictif : les questions s'enchaînent automatiquement sans intervention du joueur, qui finalement, a toujours sous les yeux une question à laquelle il peut répondre d'un clic. La modalité de jeu est tellement simple, rapide et fluide que même le joueur qui aurait décidé d'arrêter se laisse facilement entraîner à faire quelques parties de plus, et ce d'autant plus qu'il voit son score progresser entre deux questions ;

- il peut être amusant, voire accrocheur, en particulier sur les termes polysémiques (par exemple, est-ce que *chien (tarot)* peut *aboyer* ?) ;

- les parties sont très courtes : elles ne durent que quelques secondes. Cependant la séquence de jeu peut être longue (plus d'une demi-heure) de l'aveu même des joueurs ;

- les réponses données sont spontanées : est-ce mieux si les joueurs n'ont pas de compétences particulières en linguistique ? Cela génère des données de connaissance générale commune et non une connaissance de spécialistes. Notons cependant que si la question est trop complexe ou ambiguë, cela peut entraîner un grand nombre de non-réponses ;

- le résultat (comment se situe-t-on ? est-on en accord, ou au contraire, en désaccord avec la majorité des autres joueurs ?) est donné immédiatement : le joueur se sent valorisé par une réponse conforme à la réponse majoritaire (« j'ai bien répondu ») ; en cas de question plus délicate où aucune majorité nette ne se dégage, il peut se dire « je suis plus malin qu'eux », s'il a hésité et a le sentiment d'avoir fourni une réponse astucieuse ou réfléchie.

LikeIt : *LikeIt*⁴ aborde les relations de type *sentiment* simplifié, avec des questions relativement simples posées au joueur, sur le modèle *Est-ce que vous aimez l'idée de* suivi d'un terme. Les seules réponses possibles sont *oui*, *non* et *neutre*. *LikeIt* enrichit ainsi le réseau lexical concernant la *polarité* des termes. L'exploitation préliminaire de ces données seules dans le cadre d'une désambiguïsation lexicale semble montrer que la polarité prise isolément permet de sélectionner le bon usage d'un terme en contexte dans 50% des cas. Ce type de données peut aussi être exploité en analyse d'opinion.

LikeIt est un jeu de consensus dont les propriétés sont développées ci-dessous. En ce qui concerne son intérêt et ses qualités, on relève un certain nombre de caractéristiques notables :

- comme *AskIt*, *LikeIt* est un jeu addictif, à parties courtes et relance immédiate ;

⁴*LikeIt* est accessible à l'adresse <http://www.jeuxdemots.org/likeit.php>.

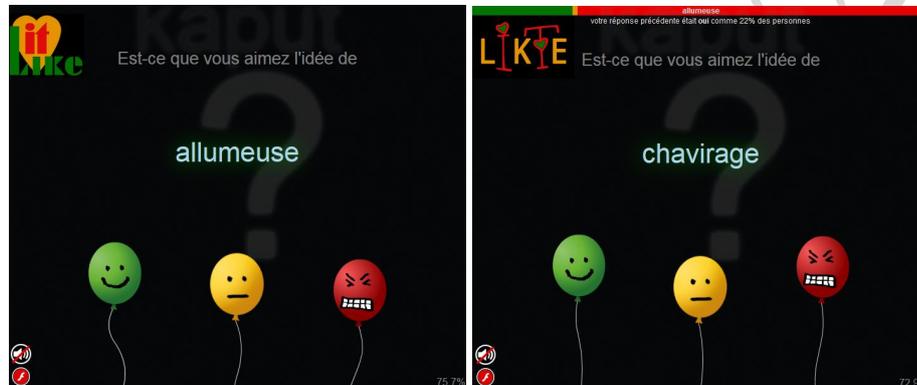


Figure 5.9.* Partie typique de LikeIt. Exemple de deux écrans consécutifs de LikeIt. Suite à la réponse donnée dans l'écran de gauche, le joueur voit immédiatement en haut de l'écran suivant (image de droite), le pourcentage de joueurs qui partagent son avis : le jeu fournit ainsi un retour direct au joueur avec une relance immédiate (nouvelle question)

- les réponses très simples : positive/négative/neutre, s'apparentent à celles requises lors d'un sondage ; mais la diversité du vocabulaire proposé et des sujets qu'il évoque est telle que, d'après les retours, les joueurs n'ont pas l'impression de répondre à une enquête d'opinion.

- certains termes suscitent des sentiments très partagés (par exemple, le terme *bloc opératoire* est positif dans l'absolu, mais négatif lorsqu'on est personnellement concerné) ; l'avis d'un même joueur peut évoluer dans le temps, en fonction des circonstances (par exemple le terme *baccalauréat* suscite un sentiment négatif quand on est lycéen, mais beaucoup plus positif lorsqu'on l'a obtenu).

La sélection d'un terme à proposer au joueur est effectuée pseudo aléatoirement dans le réseau lexical de *JeuxDeMots* selon une approche par propagation. Un terme déjà polarisé partiellement positivement ou négativement (mais pas uniquement neutre) est choisi au hasard et un des termes qui lui est relié est sélectionné. Ainsi, cet algorithme très simple (dit *de marche aléatoire*) effectue une propagation au sein du graphe entre termes potentiellement intéressants pour l'acquisition de valeurs de polarité. Une sélection strictement aléatoire au sein du réseau lexical donnerait une proportion trop importante de termes neutres ce qui diminuerait l'intérêt du jeu. En trois mois, plus de 24 000 termes ont été *polarisés* (c'est-à-dire, dotés d'une information de polarité) pour un total dépassant 150 000 votes. En trois ans, presque 150 000 termes ont été *polarisés* pour un total de 6 400 000 votes⁵.

Le jeu *LikeIt* présente un biais possible : pour un terme polysémique, il est possible que la réponse du joueur subisse une *contamination* par un sens anecdotique fortement polarisé. Par exemple, *vache*

⁵<http://www.jeuxdemots.org/likeit.php?action=list>.

Le projet JeuxDeMots

dont le sens premier, l'animal, est globalement neutre (voire légèrement positif) peut être contaminé par le sens de *vache* (*méchant*), qui a une forte polarité négative. Ainsi, les joueurs, se trouvant de facto par le jeu dans un contexte de polarité, vont voter en pensant sens le plus polarisé, et donc choisir une polarité négative pour le terme *vache*.

SexIt : Le jeu *SexIt* [Lafourcade et Fort, 2014] est très similaire à *AskIt* et vise à annoter des termes en fonction de leur relation réelle ou supposée au concept de sexe. Ce type d'information trouve des utilisations en filtrage de documents, en particulier ceux à caractère pornographique. La consigne est volontairement ambiguë, en ce que la notion de sexe peut faire référence aussi bien à la biologie et la reproduction, qu'aux comportements ou pratiques sexuelles, ou aux contenus jugés licencieux. Il est donc certain qu'indépendamment de la qualité des données recueillies, des post-traitements sont nécessaires si on souhaite distinguer les différents usages.

La première version du jeu proposait au joueur, au bout d'une vingtaine de parties, un profilage humoristique de type *complètement obsédé/oie blanche*. Ce profilage, qui n'avait aucune prétention psychologique, visait simplement à amplifier l'aspect ludique du jeu. Le calcul du profil était relativement trivial : on était qualifié d'*obsédé* (respectivement, d'*oie blanche*) si on avait une majorité de réponses de type *sex* (respectivement, de type *no sex*) pour des termes que les autres joueurs trouvaient majoritairement *no sex* (respectivement *sex*).



Figure 5.10.* Partie typique de *SexIt*. Le joueur doit choisir entre deux possibilités : est-ce que le mot a un rapport avec le sexe (cœur non barré) ou non (cœur barré). Le clic amène directement à une autre partie, avec le résultat du choix précédent (pourcentage de joueurs du même avis) en haut de l'écran.

A l'issue d'environ 5 000 parties sur environ 500 termes différents, il s'est avéré que les données étaient fortement biaisées. Plus précisément, des termes n'ayant *a priori* pas de rapport avec le concept de sexe avait en moyenne un tiers des votes en faveur d'une connotation sexuelle (*sex*). De façon symétrique, des termes clairement sexuels avaient environ un tiers de votes *no sex*. Ce

pourcentage de votes inattendu était bien trop élevé pour être attribuable à une variation naturelle due aux erreurs ou aux quelques joueurs pratiquant le trollage.

L'analyse comportementale de joueurs-testeurs a montré qu'environ un tiers d'entre eux avaient tendance à choisir leurs réponses en fonction de ce qu'ils souhaitaient obtenir comme qualificatif, à savoir *obsédé* ou *oie blanche*. Dès lors, le choix n'était plus du tout fondé sur une appréciation personnelle mais sur un objectif à atteindre *a priori*, à savoir être catalogué soit comme *obsédé sexuel*, soit comme *oie blanche*... Atteindre cet objectif était immédiat en cliquant systématiquement sur un seul des deux choix, sans même à avoir à réfléchir ni même lire le terme proposé.

La mise à jour de *SexIt*, en supprimant les mentions *oie blanche* et *obsédé* et en réinitialisant les données a permis de régler immédiatement le problème. Dans les données obtenues *via* la nouvelle version, le bruit/variabilité ne dépasse pas 5 %.

Il s'agit ici d'un exemple typique où la boucle de jeu, mal analysée à l'origine, altérait la qualité des données produites en les faussant. L'objectif fixé par le jeu suppose la création par le joueur de ressources de qualité, c'est-à-dire reflétant fidèlement sa pratique langagière. Tout autre objectif que serait tenté de se donner le joueur, dans la mesure où il est susceptible de biaiser les données, doit être neutralisé.

Un autre biais possible avec *SexIt* viendrait du fait qu'hommes et femmes ne comprennent pas la consigne de manière analogue. Les femmes auraient tendance à interpréter la consigne de façon à ne répondre oui que si le terme proposé fait invariablement penser au sexe quel que soit le contexte. Les hommes, eux, auraient tendance à répondre oui si le terme peut éventuellement avoir un rapport avec le sexe. On peut estimer que ces biais se compensent globalement.

5.6. Jeux à sélections multiples

Les jeux à sélections multiples, donnent la possibilité de répondre à une consigne en sélectionnant plusieurs propositions, voire pour certains jeux, d'en saisir *via* un champ de texte. Il s'agit donc d'une modalité de jeu intermédiaire entre *JeuxDeMots* et les jeux à votes.

ColorIt : Le jeu *ColorIt*⁶ donne la possibilité au joueur d'associer une ou plusieurs couleurs au mot proposé [Lafourcade *et al.*, 2014]. Le joueur peut également passer, ou associer une information de *non couleur* (pour les termes abstraits notamment). Le joueur peut sélectionner différentes catégories de termes : *facile/dur*, *des animaux* et/ou *des plantes*.

Le jeu incite à être efficace et exhaustif, c'est-à-dire à proposer plusieurs couleurs, bien que cela soit moins immédiat que de simplement cliquer sur une des 15 couleurs de la palette. En jouant, on accumule des points, et on franchit des niveaux. Plus le niveau est élevé, plus on peut rentrer de

⁶*ColorIt* est accessible à l'adresse <http://www.jeuxdemots.org/colorit.php>.

Le projet JeuxDeMots

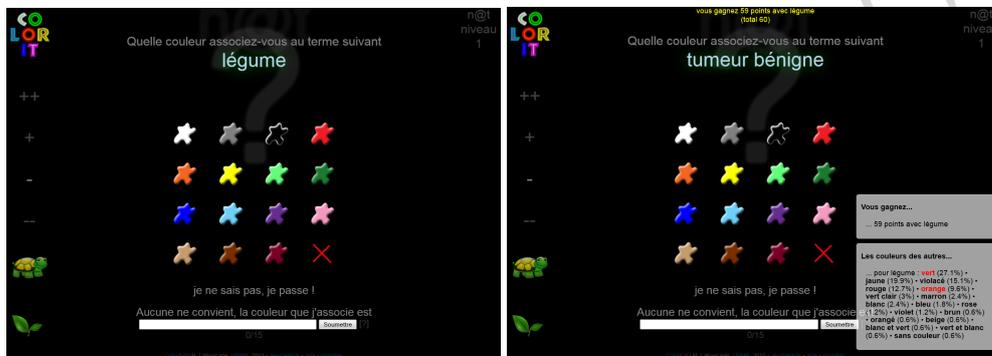


Figure 5.11.* Partie typique de ColorIt. Pour le mot légume, le joueur doit soit cliquer sur la couleur qu'il associe le plus spontanément à ce terme, soit en proposer plusieurs via le champ de texte. Sa réponse déclenche l'apparition d'un nouveau terme avec, en haut de l'écran le gain en points, et à droite les réponses des autres joueurs pour le terme précédent, avec leurs pourcentages.

caractères dans la zone de texte. Ainsi, la frustration engendrée par un niveau trop bas (on ne peut saisir qu'une couleur ou deux) incite le joueur à continuer à jouer, afin de pouvoir proposer plus de réponses. La possibilité de s'écarter de la palette proposée permet de collecter des noms de couleurs plus précis que les 15 choix possibles par défaut.

La notification du résultat affiche en pourcentage l'ensemble des couleurs qui ont été proposées par tous les joueurs pour le terme. Outre le gain en points (d'autant plus élevé que la ou les couleur(s) donnée(s) se rapprochent de celles choisies par les autres), ces données constituent une autre forme de récompense pour le joueur : non seulement il découvre les couleurs que les autres attribuent au terme, et auxquelles il n'a éventuellement pas pensé, mais il voit comment se situent ses propres réponses.

Environ 800 termes relatifs à des couleurs ou à des aspects ont été collectés au moyen de *ColorIt*. Environ 18 000 mots ont été associés à ces termes de couleurs, ce qui a généré environ 32 000 relations via 430 000 votes.

Emot : Le jeu *Emot*⁷ est basé sur le même principe que *ColorIt* et en reprend le fonctionnement et l'interface, mais cette fois-ci le joueur est invité à se prononcer sur les émotions/sentiments qu'il associe au terme cible (figure 5.12). A nouveau, l'intérêt du jeu réside dans la comparaison de ses propres réponses avec celles des autres joueurs. La possibilité de choisir entre termes faciles (courants) et difficiles (plus rares) ajoute du piment au jeu.

La sélection des termes proposés se fait automatiquement selon un algorithme dit de *propagation*, dont le principe simplifié est le suivant : un terme déjà associé à une *émotion* ou un *sentiment*

⁷Emot est accessible à l'adresse <http://www.jeuxdemots.org/emot.php>.

Le projet JeuxDeMots

est sélectionné aléatoirement dans le réseau lexical. Ensuite, l'algorithme propose, de façon équiprobable, soit ce terme, soit un terme qui lui est relié dans le réseau. Ainsi, au fur et à mesure des parties, les termes se voient associer des valeurs d'*émotion/sentiment* de proche en proche. Cette approche, à défaut de l'éviter, permet de minimiser le risque de proposer trop souvent des termes peu ou pas pertinents pour l'association avec un *sentiment* ou une *émotion*, donc peu intéressants dans le contexte de ce jeu.



Figure 5.12.* Partie classique de Emot. Le joueur doit associer au terme *générosité* un ou plusieurs émotions/sentiments, soit en cliquant sur un smiley, soit en saisissant des mots dans la zone de texte. Comme pour ColorIt, la réponse déclenche la proposition d'un nouveau terme, avec affichage du gain en points, et des réponses des autres joueurs avec leurs proportions.

Environ 1 400 termes désignant une *émotion* ou un *sentiment* ont été collectés grâce à ce jeu. Environ 7 000 mots ont été associés à ces termes d'*émotion/sentiment*, ce qui a créé environ 20 000 relations via 1 500 000 votes.

PolitIt : Le jeu *PolitIt*⁸ est relativement similaire aux deux précédents, mais cette fois-ci, les joueurs sont invités à se prononcer sur la *couleur politique* qu'ils associent au terme cible, en en sélectionnant une seule à partir d'un choix fermé (extrême gauche, écologistes, gauche modérée, centre droit, droite classique et extrême droite). Le choix déclenche la proposition d'un nouveau terme et l'affichage de trois types d'informations sur le vote précédent : le pourcentage de joueurs ayant fait le même choix, la répartition des votes des autres joueurs entre les six possibilités, et la répartition plus globale droite/gauche.

⁸ *PolitIt* est accessible à l'adresse <http://www.jeuxdemots.org/politit.php>.

Le projet JeuxDeMots

L'algorithme de choix des termes est similaire à celui utilisé dans le jeu *Emot*, à ceci près que le critère de sélection est d'être relié à une valeur désignant une *tendance politique* (et non d'*émotion* ou de *sentiment*).



Figure 5.13.* Partie classique de PolitIt. Le joueur doit associer le terme *race* avec une tendance politique. Un seul choix est possible. L'image du bas montre le pourcentage de joueurs ayant fait le même choix, et la répartition des réponses des autres joueurs entre les 6 possibilités, et entre la droite et la gauche (ici, un cas particulier montrant un choix unanime).

Tierxical, un jeu d'ordonnancement : Dans *Tierxical*⁹, le joueur est invité à classer par ordre de pertinence les trois termes qui, selon lui, répondent le mieux à la consigne. La conception et l'interface du jeu font clairement référence à un *PMU lexicalisé*, où les chevaux seraient des termes associés *via* une relation à un mot cible.

Le gain en points dépend de la correspondance entre le choix du joueur et l'état du réseau lexical. L'intérêt de ce jeu est d'ajuster les poids de relations déjà présentes dans le réseau en fonction des avis des joueurs. Un joueur qui a le *tiercé dans l'ordre* a non seulement choisi les 3 termes les plus associés (donc pour lesquels les poids des relations avec la cible sont les plus élevés), mais il les a bien classés, c'est-à-dire par ordre de poids décroissants. Le *tiercé dans l'ordre* va donc avoir pour effet de confirmer l'état du réseau, en augmentant les poids respectifs des 3 relations. Mais en cas

⁹*Tierxical* est accessible à l'adresse <http://www.jeuxdemots.org/tierce/index.php>.

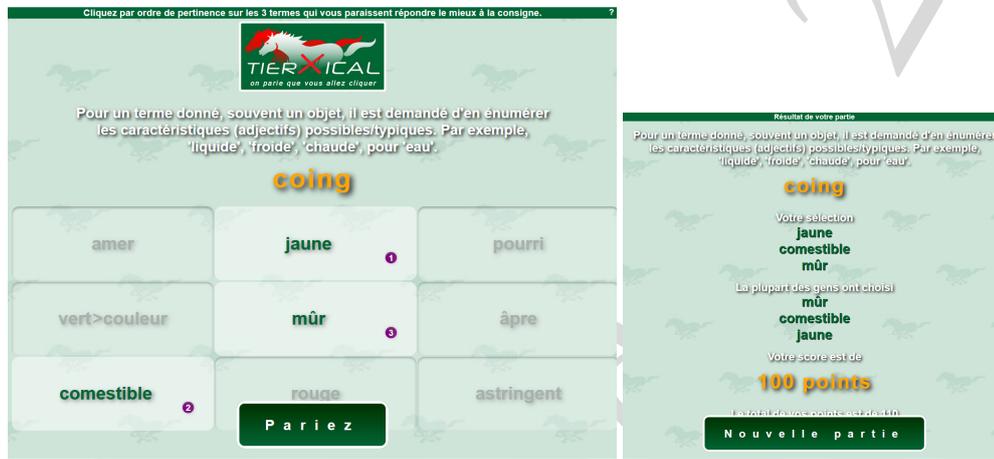


Figure 5.14.* Partie typique de Tierxical. Parmi 9 caractéristiques possibles, le joueur est invité à sélectionner dans l'ordre les trois qu'il estime les plus pertinentes pour le terme *coing*. A droite, on voit que le résultat de cette partie est un *tiércé* dans le désordre

de *tiércé* dans le désordre (le joueur a choisi les 3 termes les plus associés, mais pas dans le bon ordre), ou d'un choix ne correspondant pas à un *tiércé*, le joueur gagne moins de points, mais ses réponses ont exactement le même impact sur l'état du réseau : le premier choix augmente le poids de la relation de 5 points, le second de 3 et le troisième de 1. Le calcul et l'attribution des points n'est qu'un effet de jeu pour motiver les joueurs à faire les choix qui leur semblent les plus honnêtes. Le jeu est excitant parce qu'il est assimilable à un pari sur ce que les autres joueurs ont mis en moyenne.

Le joueur a également la possibilité d'émettre des choix *négatifs*, en cliquant de nouveau sur un terme choisi, s'il estime que certaines propositions ne conviennent pas. Le joueur entre ainsi dans un mode de jeu avancé, où il s'agit de sélectionner les trois propositions convenant le moins. Cette possibilité permet soit d'introduire, soit de renforcer des poids négatifs pour certaines relations (en indiquant, par exemple, qu'une *autruche* ne vole pas, ou que la *glace* n'est pas *chaude*).

Depuis le lancement du jeu à la mi-2012, environ 35 000 parties ont été jouées et ont ainsi participé activement à l'ajustement des poids des relations dans le réseau lexical. Aucune partie n'est mémorisée, la comparaison entre les réponses du joueur et l'état du réseau se fait de façon synchrone.

5.7. Des jeux vers les systèmes contributifs

L'expérience du projet *JeuxDeMots* a montré que certains joueurs très investis, ayant développé une bonne connaissance de tous les rouages de *JeuxDeMots* manifestaient intérêt et curiosité vis-

Le projet *JeuxDeMots*

à-vis du projet en lui-même, et désiraient contribuer de façon plus directe à la création du réseau lexical. Ces joueurs ne cessent pas de jouer pour autant, mais leur intérêt prononcé pour le langage et le vocabulaire les pousse à tenter de comprendre comment le jeu construit le réseau et à s'y investir autrement. La fierté de participer à un projet collectif, à la constitution d'une ressource lexicale librement accessible devient alors, pour ces joueurs, un élément de motivation qui prime sur l'attrait du jeu en tant que simple divertissement.

Asku, un pseudo-jeu d'apprentissage : Dans le pseudo-jeu *Asku*¹⁰, un mot est présenté au joueur, ainsi qu'une série de questions sur les relations que ce terme entretient avec d'autres mots au sein du réseau. Pour chaque affirmation, la réponse du joueur, qui va devoir choisir entre *vrai/possible/peu pertinent/bof non/faux* permet au système de déduire d'autres connaissances, qui feront à leur tour l'objet de questions. Le but est d'essayer de faire apprendre (par inférence) au système un maximum de choses. Le joueur voit immédiatement l'effet de ses réponses : en bas de la fenêtre de jeu, le système affiche combien de choses il a apprises (c'est-à-dire, combien de déductions il a réalisées) suite à la réponse, et 3 barres de progression cumulent les réponses en termes qualitatifs : apprentissage (quand la réponse amène une information nouvelle pour le système), connivence (quand la réponse confirme une information), et contrariété (quand la réponse est contradictoire avec une connaissance déjà présente dans le système). De plus, le joueur est invité à collaborer à l'enrichissement du réseau de façon ouverte en entrant des informations sur diverses relations du terme cible *via* un champ de texte libre.

Diko, un environnement contributif : *Diko*¹¹ est une interface Web de consultation et de contribution associée au réseau lexical de *JeuxDeMots*. L'utilisateur peut consulter les termes associés par diverses relations au terme de son choix *via* un formulaire. Pour chaque relation renseignée, les termes associés sont listés par ordre de poids décroissant.

L'interface de *Diko* permet d'entrer dans un mode *contributif*. Chaque entrée de *Diko* correspond à un terme, auquel l'utilisateur peut associer d'autres termes, pour environ 80 relations. Chaque terme ainsi associé *via* une relation constitue une *contribution*. Les contributions sont visibles par les autres contributeurs, qui peuvent voter pour ou contre. Des administrateurs pourront par la suite valider ou invalider les contributions en attente. Le principe directeur de la conception de *Diko* est de placer les utilisateurs/contributeurs dans une situation de négociation avec les autres utilisateurs. Pour chaque contribution, un *fil de discussion* peut être ouvert, soit par l'auteur de la contribution, soit par n'importe quel autre participant pour exprimer un doute, un désaccord, poser une question, justifier un vote négatif, etc.. Cette approche par négociation est à l'opposé de celle des jeux, où les joueurs expriment leur choix de façon indépendante et où l'entrée de nouvelles données dans

¹⁰*Asku* est accessible à l'adresse <http://www.jeuxdemots.org/askyou.php>.

¹¹*Diko* est accessible à l'adresse <http://www.jeuxdemots.org/diko.php>.

ASK YOU

- +

a A

11 07 07 2. elements trouve

félicité (bonheur)

| | vrai | possible | peu pertinent | bof non | faux |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| félicité (bonheur) peut faire penser à contemplation | <input type="checkbox"/> |
| félicité (bonheur) peut faire penser à approuvé | <input type="checkbox"/> |
| félicité (bonheur) peut faire penser à vision | <input type="checkbox"/> |
| félicité (bonheur) a comme synonyme bien-être | <input type="checkbox"/> |
| félicité (bonheur) peut se trouver dans le lieu lit | <input type="checkbox"/> |
| félicité (bonheur) peut faire penser à béatitude | <input type="checkbox"/> |
| félicité (bonheur) peut faire penser à plaisir | <input type="checkbox"/> |

... ou encore que félicité (bonheur) peut faire penser à

Un autre mot !

... ou encore : félicité

... ou : sémiologie • psammophile • nager le crawl • tisser une toile • PAO (informatique)

Figure 5.15.* Ecran de AskU. Le joueur/contributeur est invité à s'exprimer sur un certain nombre de propositions concernant le terme *félicité* dans le sens de bonheur. Pour chaque proposition, il dispose d'une gradation de réponses possibles, allant de l'assentiment sans réserve (vrai) au rejet total (faux).

le réseau est le résultat d'une confrontation effectuée par le système (on rappelle que pour qu'une relation entre dans le réseau *via JDM* il faut que le terme correspondant ait été proposé par les deux joueurs dont les réponses sont comparées au cours d'une partie). Dans l'approche contributive de *Diko*, l'entrée d'une nouvelle relation est dite *négociée* : ce sont les contributeurs qui débattent et votent, et un administrateur humain qui tranche.

L'interface de *Diko* est extrêmement peu ludifiée, la seule concession à la dimension ludique étant l'affichage d'un *Hall of Fame* classant les contributeurs en fonction de l'importance de leur participation. Être à l'origine d'une contribution validée rapporte 3 points tandis que voter pour une telle contribution procure 1 point. Plus une contribution dispose de votes positifs, plus elle a de chances d'être validée et donc de rapporter des points à son auteur. Ce mécanisme pousse les contributeurs à être à la fois prolixes et coopératifs. En effet, plus une contribution semblera pertinente plus elle recueillera d'avis positifs (votes *pour*). Jouer en solitaire sans voter pour les contributions des autres n'est pas productif, pas plus que n'être que suiveur. Il est à noter que pour les contributeurs, l'existence du *Hall of Fame* en tant que *classement* n'a pas une importance capitale, la motivation première restant l'intérêt porté au projet et à la chose lexicale.

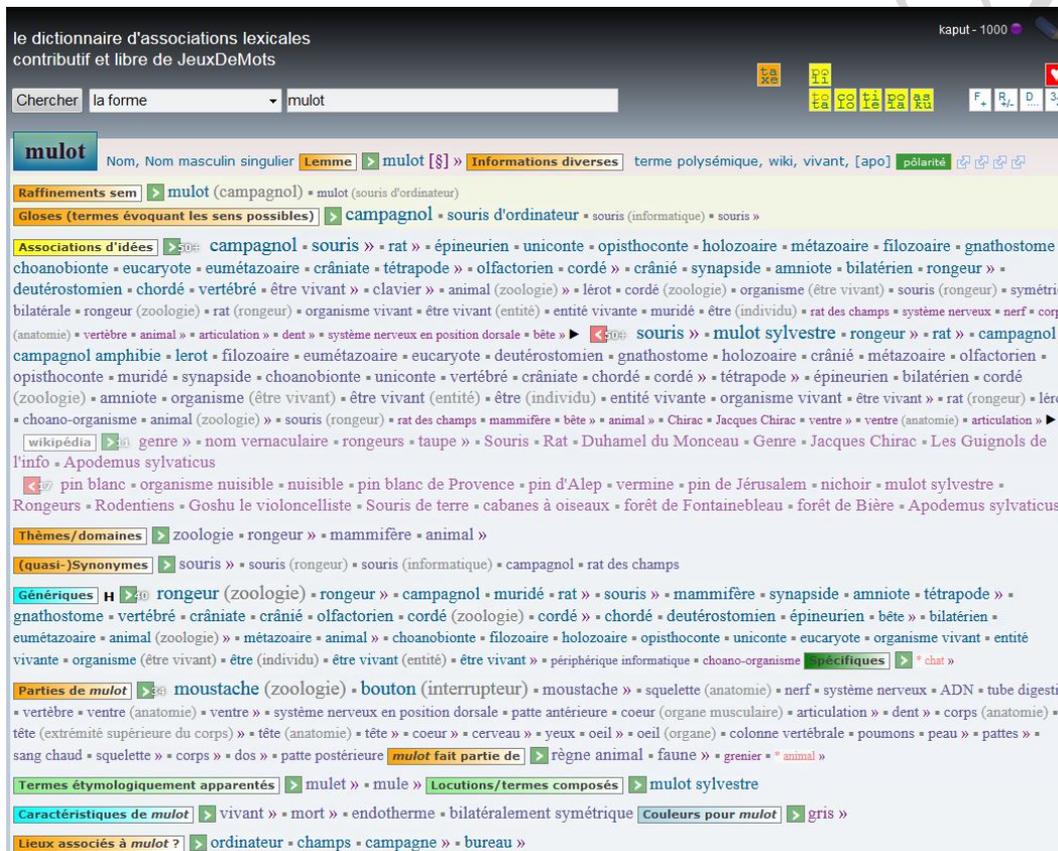


Figure 5.16.* Ecran typique de Diko. L'utilisateur a affiché l'entrée mulot, et on peut voir les termes reliés à mulot par diverses relations. L'interface est configurable et l'affichage du niveau de détails (nombre de mots à afficher, précision des informations, etc.) peut être ajusté par l'utilisateur

Les joueurs/contributeurs les plus actifs ont ainsi gagné plusieurs centaines de milliers de points. On estime qu'environ 400 000 relations auraient été introduites *via Diko*. Il est à remarquer qu'il s'agit la plupart du temps de contributions particulièrement intéressantes pour la richesse du réseau, qui, de plus, auraient été difficiles à obtenir de façon non négociée *via* les jeux.

5.8. Données collectées et propriétés des jeux présentés

En guise de conclusion à ce chapitre, nous évoquons les données collectées, ainsi que quelques propriétés des jeux que nous avons présentés. En analysant les données du réseau *JeuxDeMots*, on décèle facilement les difficultés qu'éprouvent les joueurs face à certains types de relations lexico-



Figure 5.17.* Ecran de Diko en mode contribution. L'utilisateur a affiché l'entrée grande marée, et on peut voir les termes proposés mais non encore validés (en vert). Les carrés rouges et verts correspondent à des possibilités de vote pour ou contre la contribution. Les petits cœurs correspondent à la polarité. L'interface est configurable et les couleurs des étiquettes des relations ont été ajustées par l'utilisateur.

sémantiques. Soulignons qu'ils évitent naturellement de jouer les relations qu'ils trouvent trop difficiles, ou inintéressantes. On note le même comportement quand il s'agit de contribuer à la ressource via l'interface *Diko*.

Depuis le lancement du projet *JeuxDeMots*, le réseau lexical a acquis plus de 12 millions de relations. La base, initialement de 150 000 termes, dépasse actuellement 420 000 (au 1^{er} octobre 2014). Il y a environ 60 relations lexico-sémantiques dans JDM, la plupart d'entre elles pouvant faire l'objet de parties de JDM, ou de questions dans les contre-jeux.

Quelques chiffres

Certaines des relations du réseau lexical ne sont pas jouables, notamment :

- - 52 0148 catégories morphosyntaxiques : jouer sur les parties du discours serait sans intérêt ;
- 78 312 lemmes, obtenus de façon automatique ou contributive (hors jeux).

Le projet *JeuxDeMots*

Ces informations ne peuvent être acquises *via* des parties parce qu'elles sont trop peu variables pour présenter un intérêt ludique. Il est donc nécessaire de les entrer manuellement ou par tout autre moyen (semi) automatique.

- 22 842 raffinements sémantiques issus du calcul de cliques, dont l'approche a été présentée dans [Lafourcade et Joubert, 2010] et/ou des contributions des joueurs.

Les joueurs ne rentrent pas directement les raffinements sémantiques mais ont le loisir de jouer à *gloser* les termes, c'est-à-dire de proposer *via* une partie des termes représentatifs des usages possibles. Les *raffinements* seront majoritairement choisis parmi ces *gloses*. Les joueurs trouvent la consigne *Donner des sens/significations pour ce terme* particulièrement difficile à comprendre, mais intéressante. Parfois, l'introduction de périphrases semble être la seule réponse possible si aucun hyperonyme ou synonyme ne semble convenir.

Des consignes/rerelations difficiles

Enfin, certaines relations posent plus de problèmes aux joueurs :

- cause/conséquence et producteur/produit

Comme en témoignent certains fils de discussions en marge du jeu, ou certains débats lors de procès, il semble que les joueurs aient souvent du mal à discriminer clairement la cause et la conséquence, ou encore le producteur et le produit. Manifestement, la confusion n'est pas d'ordre linguistique, mais plutôt en relation avec la connaissance du monde, la perception personnelle de l'ordre des choses. Ainsi, la relation *cause/conséquence* (ou *conséquence/cause*) semble difficile à établir entre les termes *fièvre* et *malade*.

De façon croisée, il est parfois difficile de choisir entre *producteur* et *cause*, d'une part et *produit* et *conséquence* d'autre part.

- la différence entre *instance* et *hyponyme*

Pour beaucoup de joueurs, les hyponymes semblent inclure les *instances* (mais pas l'inverse). Par exemple, *Jolly Jumper* est vu aussi bien comme une *instance* de *cheval* que comme un *hyponyme*. Par contre, *pur sang* n'est (à juste titre) considéré que comme un hyponyme.

- la différence entre le *rôle agentif* et les étapes de construction

Le *rôle agentif* fait référence à la manière dont un objet est créé/construit. A ce titre, ce sont des verbes de construction de l'entité/artéfact qui sont requis, par exemple, *construire* pour *maison*, ou *écrire/éditer* pour *livre* [Pustejovsky, 1993]. L'implication agentive concerne le ou les verbes impliqués dans le processus de construction de l'objet (par exemple pour un *gâteau* : *mélanger*, *enfourner*, *cuire*, *démouler*, etc.). La distinction entre la sémantique de chacune des deux relations n'est pas immédiatement intuitive pour les joueurs et il s'ensuit un nombre important de parties erronées.

Le projet JeuxDeMots

- la distinction entre *magn* (plus intense) et bon (mélioratif).

L'axe d'intensité relatif à un terme est manifestement parfaitement subjectif pour les joueurs, qui souvent, à côté de réponses qui respectent parfaitement la consigne *est plus intense que* (dont un exemple canonique est *forte fièvre* ou *fièvre de cheval* comme *magn de fièvre*), indiquent des termes qui sont plutôt d'ordre mélioratif (*excellent repas* est davantage un mélioratif qu'un *magn de repas*).

- les sentiments/émotions

La notion de sentiment/émotion semble parfois très extensible, et il arrive que certains joueurs attribuent la valeur de sentiment ou d'émotion à des termes qui relèvent plutôt de l'association d'idée. Par exemple, on va trouver *richesse* comme sentiment/émotion associé à *or* (*argent*). Le terme *richesse* porte alors le sens de *sentiment de richesse*.

- termes de la même famille lexicale

Initialement, l'objectif était d'obtenir, *via* les parties sur cette relation, des mots morphologiquement dérivés du terme cible. La consigne *donner des mots de la même famille pour le terme qui suit* a été interprétée par une majorité de joueurs aussi bien dans le sens de famille *étymologique* que de famille *morphologique*, les gens ne faisant pas une distinction claire entre les deux notions. (Notons que quelques désaccords sur cette question ont donné lieu à de vifs débats dans le cadre des procès ou de fils de discussion, et ont eu le mérite de conférer une certaine expertise à quelques joueurs passionnés.) Les résultats, parfois étonnants, témoignent non seulement de cet amalgame entre étymologie et morphologie, mais aussi de confusions liées à l'homophonie. Ainsi, pour le substantif *nuit*, nous obtenons : *nocturne, noctambule, noctambulisme, noctuelle, noctule* qui sont étymologiquement apparentés, *nuitée, nuitamment, minuit* qui sont morphologiquement dérivés, mais aussi des intrus comme *nuisible, nuire*, qui ont une partie homophone avec *nuit*, et enfin *chemise de nuit, bonnet de nuit*, qui ne sont que des mots composés incluant le mot cible. Un échantillonnage évalué à environ 5% les associations erronées concernant cette relation (les erreurs étant parfois largement répandues parmi les locuteurs). Pour résumer, sont validés comme *termes de la même famille* les mots étymologiquement apparentés et les mots morphologiquement dérivés, la distinction entre les deux types ayant lieu en aval, de façon automatisée.

- problème avec la typicalité

Pour certaines relations (comme *agent, patient, caractéristique*, etc.) nous nous attendions *a priori* à obtenir des associations *typiques*. Par exemple *rugir* est une action *typique* pour un *lion*, tandis que *courir* est *possible*, mais pas spécialement *typique*. De même, *tigre* est typiquement *féroce*, mais peut aussi être *vieux*. Nous constatons à la vue des données recueillies que l'ensemble des possibles a tendance à apparaître dans le réseau, mais avec des pondérations beaucoup moins fortes que celles des associations typiques. C'est une bonne nouvelle puisque là où nous ne nous at-

tendions à récolter que l'information *typique*, nous récoltons également l'information *possible*, donc des données supplémentaires, avec des pondérations permettant une distinction nette.

Forçage, typologie de joueurs et taux d'erreur

Faudrait-il pénaliser un joueur qui fournirait une réponse erronée ? Nous ne le pensons pas. Tout d'abord, cela supposerait une définition précise de ce qu'est une relation erronée, assortie d'un mode de détection efficace et sans ambiguïté. De plus, le risque d'être pénalisé peut être facilement inhibant, voire carrément dissuasif en cas d'hésitation, tandis que des pénalisations répétées peuvent entraîner une démotivation et une désaffection pour le jeu, les deux effets allant à l'encontre du but recherché.

Un des biais majeurs de JDM est une tendance globale des joueurs à « forcer la consigne », c'est-à-dire à vouloir répondre coûte que coûte – « il faut absolument trouver et mettre des réponses ». Bien que le joueur ait la possibilité, soit d'indiquer qu'il n'existe pas de réponse compatible avec le mot et la consigne, soit de passer s'il n'est pas inspiré, il arrive qu'il préfère quand même indiquer ce qui lui semble le plus approchant. C'est ainsi par exemple que pour des spécificités de *semaine* ont été proposés les jours de la semaine : lundi, mardi, etc. Les relations erronées relatives à ce phénomène de *forçage* ont été évaluées (par échantillonnage) à environ 0,1 % des relations produites. Il est à noter que la présence des contre-jeux vient largement contrebalancer ce biais propre aux jeux ouverts.

Ceci étant dit, nous avons remarqué que les espaces d'échanges entre joueurs (les procès, le forum, et dans une moindre mesure, la fenêtre de messagerie instantanée) sont des lieux privilégiés pour éclaircir la sémantique des relations quand elle n'est pas bien comprise par les joueurs. Un joueur assidu va ainsi voir son expertise augmenter nettement au fil du temps.

Nous avons évalué plus précisément les performances globales des joueurs. Dans JDM, chaque partie en stock est « signée » (nous savons quel joueur en est le créateur) et on peut calculer de façon incrémentale le nombre moyen de réponses communes entre les joueurs deux à deux (on parlera d'*efficacité*). De plus, grâce au concours de joueurs volontaires, nous avons une idée partielle de leur profil (âge, sexe, niveau d'étude, domaines de compétence, etc.). Ceci nous a permis de mettre en évidence les faits qui suivent.

L'efficacité moyenne des joueurs est de 2,7 (c'est-à-dire qu'ils ont en moyenne 2,7 termes en commun). Les meilleurs joueurs ont une efficacité aux alentours de 5 voire 6. Les joueurs dont l'activité professionnelle relève du domaine de la linguistique au sens large (phonétique, linguistique, informatique-linguistique, gestion documentaire, etc.) ont une efficacité de 1,8 avec l'ensemble des autres joueurs et de 2,2 entre eux. L'analyse de leurs parties montre que les *linguistes* font moins de propositions que la majorité des autres joueurs, et que leurs propositions sont majoritairement parmi les termes les plus activés du réseau. Donc, la plupart du temps, il s'agit de propositions correctes

Le projet *JeuxDeMots*

mais peu originales et n'alimentant pas la longue traîne¹². Les joueurs ayant une activité professionnelle relevant de la langue (linguistes, FLE, TAL, etc.) représentent environ 8 % des joueurs, et 6 % de l'activité.

Une évaluation statistique globale a permis d'estimer à 0,5 % le nombre de relations manifestement erronées, et à moins de 2 % celles qui sont discutables. Environ 0,4 % des relations erronées sont repérables automatiquement. Il s'agit par exemple des doubles hyperonymes (A hyperonyme de B et B hyperonyme de A) qui modulo la polysémie, posent problème. Les faux négatifs sont estimés à environ 40 %. Le 0,1 % restant doit être débusqué au cas par cas mais concerne des relations de faible poids et tend à se résorber naturellement avec les contre-jeux. L'expérience de *Totaki* a montré que sur un jeu de devinettes et strictement à partir du réseau de *JeuxDeMots*, le système trouvait la bonne réponse dans 75 % des cas (à comparer à 48 % pour des interlocuteurs humains).

En conclusion, il nous semble important de souligner deux choses. L'ensemble des jeux du portail JDM vise à construire une ressource par consensus. Cela signifie que contrairement à un jeu de questions classique, comme un quizz par exemple, il n'y a pas de bonnes (justes) ou de mauvaises (fausses) réponses. Ce qui définit une réponse comme valide, c'est le nombre de joueurs qui la considèrent comme telle, puisqu'ils l'ont fournie. Ainsi, le gain de points n'est fonction que de l'adéquation entre une réponse donnée et l'ensemble des réponses des autres joueurs à la même question.

Enfin, les données du réseau lexical JDM, qui résultent en grande partie de l'activité des joueurs, sont en accès libre et libres de droits. Elles peuvent être récupérées globalement sur la base d'une exportation mensuelle ou dynamiquement terme par terme.

¹²http://fr.wikipedia.org/wiki/Longue_tra%C3%Aene.

Le projet *JeuxDeMots*

Version auteurs, dec. 2014

Conclusion

Qualités et avantages d'un bon GWAP ?

Au vu de notre expérience sur le projet *JeuxDeMots*, ainsi que de l'analyse d'un certain nombre d'autres jeux, nous allons tenter de définir les caractéristiques nécessaires pour qu'un GWAP ait du succès. Celles-ci peuvent se trouver dans les réponses à quelques questions simples : que veut-on demander aux joueurs ? que peut-on leur demander ? sous quelle forme ? La quasi-totalité des éléments de la grille d'analyse qui suit, bâtie à partir des jeux du projet *JDM*, est transposable à des jeux non lexicaux. Peu d'analyses ont été proposées concernant les GWAP et la plupart se situe dans un contexte très général [von Ahn et Dabbish, 2008] ou constitue plus une description de ce qui existe [Thaler *et al.*, 2011].

... pour les joueurs

Pour attirer des joueurs, puis les inciter à revenir participer, un jeu doit posséder un certain nombre de qualités (liste bien évidemment non exhaustive) :

- amusant

Le jeu doit présenter un intérêt ludique (n'est-ce pas une évidence ?) au niveau de l'interface afin d'accrocher les joueurs, mais plus encore au niveau du contenu afin de les garder. Le contenu doit pouvoir se renouveler avec une petite dose de répétition offrant la possibilité de s'amender sur une partie précédemment ratée. Dans l'optique de construire des ressources lexicales et terminologiques, la grande taille du vocabulaire offre des possibilités de renouvellement quasiment illimitées. Ceci est également vrai dans d'autres domaines, comme nous l'avons vu dans un certain nombre de GWAP : par exemple, le nombre de personnages que l'on peut tenter de faire deviner à *Akinator* est potentiellement illimité.

Conclusion

- facile à appréhender

Beaucoup de joueurs potentiels ne lisent pas les consignes, il faut donc que le jeu soit intuitif et que l'apprentissage soit le plus réduit possible. Dans le cadre d'un jeu lexical, de nombreuses informations et/ou consignes sont directement compréhensibles par le locuteur et ne nécessitent généralement pas un bagage linguistique élaboré. De nombreux autres GWAP ne nécessitent aucune formation et sont immédiatement accessibles : par exemple, il suffit de connaître Tetris pour pouvoir jouer à *Phylo*, aucun apprentissage n'est nécessaire pour retrouver des oiseaux et leurs œufs dans le projet *Nightjar*.

- relance par *reclie* immédiat

Outre l'intérêt intrinsèque du jeu, les joueurs jouent d'autant plus longtemps que l'accès au jeu est immédiat et sa relance instantanée. L'acquisition des ressources s'en trouve quantitativement renforcée. Minimiser le délai et le nombre de clics nécessaires pour lancer une nouvelle partie est une stratégie à privilégier autant que possible.

- valorisation des joueurs

Un système de points et de classement entre les joueurs rend en général le jeu plus attirant : instaurer et stimuler une compétition entre les joueurs pousse certains à y consacrer beaucoup de temps, acquérant ainsi pratique et expérience, ce qui les rend non seulement plus productifs mais augmente globalement la qualité des contributions. Il est très rare qu'un joueur soit rebuté par la présence d'un classement. Au pire, il peut choisir de l'ignorer. De plus, dans des jeux impliquant une confrontation avec le système, les joueurs peuvent être mis en situation de se sentir *plus intelligents que la machine* ; par exemple, dans *Totaki* [Joubert *et al.*, 2011] du projet *JDM*, un grand nombre de parties visent à piéger le système, ou à lui apprendre quelque chose. Les joueurs ont alors le sentiment (légitime) d'être utiles dans le processus d'acquisition des connaissances. Les concepteurs jouent souvent sur la corde sensible de la *science citoyenne*, comme par exemple dans *Worm Watch Lab* où, avouons-le, le fait de regarder des vers se tortiller mollement n'a rien de particulièrement amusant : mais il est motivant de sentir qu'on aide la science et la recherche, *a fortiori* quand l'aspect *recherche médicale* est mis en avant.

- jeu évolutif

Les joueurs acquièrent de l'expérience en jouant, ils répondent de manière de plus en plus pertinente, à des questions de plus en plus précises, de plus en plus spécialisées. Par exemple, une partie *idées associées* au terme *chat* est intéressante pour découvrir *JeuxDeMots*, mais par la suite, les joueurs préfèrent des consignes plus difficiles, faute de quoi ils risquent de se lasser. Il est nécessaire que la difficulté croisse avec l'expérience du joueur. Il est également possible de définir des niveaux de difficulté (par exemple, débutant, normal, confirmé, expert) correspondant dans le cas d'un jeu lexical à la rareté des mots proposés ou à la difficulté des relations lexicales demandées.

Conclusion

- parties courtes

Plus les parties sont courtes, moins le joueur potentiel hésite à jouer... le nombre de parties jouées sur une session est strictement contravariant avec la longueur moyenne des parties. Cette propriété implique une granularité faible des ressources que l'on cherche à récolter, et concerne donc plutôt des jeux de type microtâche. Les approches lexicales en général donnent lieu à des microtâches dont la granularité est très fine et le temps de réalisation très faible. Dans le cas de *JeuxDeMots*, une enquête a montré que les joueurs qui prévoyaient de jouer pendant une dizaine de minutes passaient en moyenne plutôt 25 minutes sur le jeu.

- format et portabilité du jeu

Si un jeu présente une interface suffisamment simple pour être jouable sur smartphone, il est possible de jouer n'importe où, et presque n'importe quand (file d'attente, transports en commun, etc.). Les joueurs jouent plus souvent et il peut y avoir éventuellement un phénomène de mode *via* du bouche-à-oreille ; ce fut notamment le cas lors du lancement d'*Akinator* sur smartphones.

... pour les concepteurs

Lorsque le but poursuivi est la constitution d'une ressource de qualité, cette méthode d'acquisition présente un certain nombre d'avantages :

- les ressources produites sont le fruit de résultats consensuels, partagés par un grand nombre (voire un très grand nombre) de joueurs ;

- le résultat est la constitution d'une base de connaissances générales communes, auxquelles viennent nécessairement s'ajouter des connaissances plus spécialisées ;

- le prix de revient des ressources est très faible : c'est celui du développement, puis de la maintenance et de l'animation du jeu ; signalons que ce dernier point n'est pas à négliger, sous peine d'observer rapidement une certaine désaffection pour le jeu ;

- en fonction du nombre de joueurs et de leur activité, l'acquisition peut être relativement rapide (par exemple, de l'ordre de plus de 1 000 occurrences de relations lexicales par jour dans le cadre du projet *JDM*) ;

- un jeu d'acquisition de ressources doit s'adresser à une partie aussi large que possible de la population. Cependant, la nature même des objets manipulés peut réduire fortement cette espérance. Ainsi, les compétences requises pour jouer à *Dizeez* (une connaissance fine et approfondie sur les associations gènes/maladies) restreignent fortement le nombre de joueurs potentiels. Inversement, *JDM* s'adresse à un public beaucoup plus large et varié : 70% des joueurs ont entre 30 et 60 ans, 60% sont des femmes, et la quasi-totalité a un niveau d'étude supérieur ou égal à bac+2 ;

Conclusion

- d'un point de vue éthique, on remarquera enfin que la sollicitation de contributeurs à travers des jeux semble ne pas poser autant de problèmes que dans un cadre rémunéré [Sagot *et al.*, 2011], même si à première vue, les tâches peuvent sembler très similaires. De plus, si le jeu est bien conçu, la motivation des gens à bien jouer est une garantie de qualité de la ressource produite, ce qui est loin d'être assuré dans le cadre d'une tâche (mal) rémunérée. Enfin, la motivation est d'autant plus grande si le jeu *bien conçu* l'est au service d'un projet intéressant et produit des données libres et accessibles.

Roulement des joueurs

Le roulement (renouvellement des joueurs) peut être relativement court, et cela d'autant plus en cas de phénomène de mode au sein d'un sous-groupe ou d'une petite communauté de joueurs. Selon notre expérience sur *JDM*, le roulement est d'environ trois semaines. Nous avons mesuré cette durée de façon continue après chaque campagne d'information, que ce soit lors de conférences scientifiques ou auprès d'étudiants. Un petit nombre de joueurs continuent de jouer pendant une longue durée (plusieurs années), d'autres découvrent le jeu, font quelques parties et ne reviennent plus. (Sont-ce des joueurs qui sont *instables* par nature, en quête permanente d'expériences nouvelles ? ou déçus par CE jeu ? dans ce dernier cas, pourquoi ?)

Nous avons également constaté qu'un certain nombre de joueurs reviennent sur le jeu, après un longue période d'inactivité, parfois simplement pour jouer à nouveau, mais le plus souvent avec une attitude différente, plus ciblée ; par exemple, ils jouent à *Totaki* pour vérifier si le système connaît telle ou telle information.

Un roulement des joueurs relativement court est plutôt un inconvénient : sur une brève période, on n'a pas le temps, d'une part d'acquérir suffisamment d'expérience et de pratique pour se voir progresser, d'autre part de découvrir et expérimenter toutes les subtilités et tous les modes de jeu, et on reste trop souvent à un niveau sommaire. Indiquons toutefois qu'il existe une proportion non négligeable de joueurs au comportement addictif, entre lesquels le système de points et de classement suscite une émulation productive : leur volonté de jouer le mieux possible et de se surpasser les pousse à se documenter, et permet d'acquérir des ressources qui s'apparentent plus à une connaissance de spécialistes.

Cependant, le roulement permet à de nouveaux joueurs, novices par définition, d'intervenir dans le jeu. Ceux-ci perpétuent la fourniture d'informations naïves (par opposition aux ressources de spécialistes). Ils permettent également de mieux prendre en considération l'évolution souvent rapide des idées, en particulier celles fortement liées à l'actualité, en fournissant des associations inédites, liées au contexte du moment.

Conclusion

Tricherie et vandalisme

Nous avons constaté deux sortes de tricherie. En premier lieu, celle qui permet de contourner certaines règles contraignantes du jeu (par exemple, la limitation du temps dans *JDM*) : ce type de tricherie ne remet pas en cause la qualité des ressources produites, voire les améliore, puisqu'en permettant de contourner des contraintes, elle rend les choses plus faciles. Cependant, elle est néfaste dans la mesure où, s'il devient évident que certains joueurs trichent, cela induit un certain écœurement chez les autres, pouvant entraîner une désaffection pour le jeu.

Pour parer cet inconvénient, nous tentons de débusquer les zones *sensibles* à l'aide de joueurs pilotes, de préférence compétents informatiquement : en effet, ce genre de tricherie est le fait de joueurs qui s'y connaissent suffisamment en informatique pour exploiter les failles du jeu en intervenant sur le code-source. D'ailleurs, dans le cadre de *JDM*, de nombreuses attaques, par injection de codes informatiques notamment, ont eu lieu. En général, ce sont les premières heures d'un tel projet qui constituent la phase critique pour ce type de risque.

Le second type de tricherie, plus grave, vise à « pourrir » la base (c'est-à-dire faire rentrer sciemment des données erronées) et relève davantage du vandalisme que de la volonté de gagner par des moyens détournés. Il s'agit pour les concepteurs d'en minimiser le risque à tout prix dès la conception, car cette tricherie-là peut irrémédiablement compromettre la qualité de la ressource que l'on cherche à constituer : fastidieuse, voire coûteuse à détecter et neutraliser manuellement (par un expert), elle est extrêmement difficile, voire quasi impossible, à détecter automatiquement. En effet, en admettant que le système soit capable de détecter une information incongrue (ce qui est déjà loin d'être évident et pose l'insoluble question des critères), la cataloguer systématiquement comme *fausse* et l'éliminer serait contre-productif : *incongru* ne veut pas forcément dire *faux* et se priver de ce type d'information inédite, mais potentiellement pertinente, va à l'encontre du but poursuivi qui est d'enrichir le système en lui fournissant des connaissances qu'il n'a pas. Si un tel système était capable de s'auto-contrôler en triant les données qu'on lui fournit, alors le projet n'aurait pas de sens.

Dans de nombreux GWAP, l'acquisition des données repose sur le principe selon lequel elles sont considérées comme valides dès lors qu'elles ont été proposées simultanément et indépendamment par au moins deux joueurs. C'est pourquoi les parties confrontent de manière anonyme et asynchrone les réponses de deux joueurs à une consigne, et le système ne mémorise que leurs réponses communes (c'est-à-dire l'intersection des deux jeux de données). Ainsi, si l'un des joueurs tente de saboter le jeu en répondant des inepties, d'une part elles ne rentreront pas dans la base de données, d'autre part l'affichage du résultat, en mettant en évidence l'échec de la tentative (aucune des inepties n'est retenue), découragera le joueur malveillant. Et si une activité suffisante est maintenue au sein du jeu, la probabilité que deux joueurs vandales se rencontrent reste faible.

Conclusion

Aspect social : le groupe des joueurs

A une époque où les réseaux sociaux sont devenus un incontournable fait de société, l'aspect communautaire, le sentiment d'appartenir à un groupe, tient souvent une large part dans l'intérêt suscité par un jeu, et les GWAP ne font pas exception.

Pour la pérennité d'un jeu, l'existence d'une communauté de joueurs, au sein de laquelle existent des interactions, présente plusieurs avantages :

- à jouer tout seul dans son coin, le joueur se lasse vite ; pouvoir visualiser *via* l'interface que d'autres personnes jouent simultanément, même si on ne les connaît pas réellement, entraîne une certaine émulation ;

- fournir aux joueurs, *via* un *forum* la possibilité de discuter entre eux, d'échanger leurs points de vue, de contester (éventuellement d'approuver) les opinions des autres, non seulement soude le groupe en favorisant la convivialité, mais rend les joueurs de plus en plus experts. De plus, tout moyen de communication mis à leur disposition leur donne l'opportunité de fournir un retour sur le jeu et ses modalités, ce qui est important pour les concepteurs ;

- il peut y avoir une forme d'appropriation du jeu par les joueurs. Par exemple, un joueur peut décider de privilégier un thème qui l'intéresse particulièrement (« je suis le spécialiste de ce domaine ») et mettre un point d'honneur à fournir des données exhaustives et minutieusement contrôlées. Cette forme de spécialisation peut faire école *via* les interactions au sein du groupe, et amener d'autres joueurs à s'estimer également spécialistes d'un domaine. A côté des connaissances générales, il s'ensuivra un enrichissement du réseau en connaissances dites *de spécialité*, et ce, dans divers domaines.

Derniers mots

Dans ce livre, nous avons montré comment les GWAP, ou jeux à finalité, permettent de résoudre certains problèmes ou d'acquérir des ressources de qualité, facilement et à moindre coût. Toutefois, cela suppose qu'ils remplissent un certain nombre de conditions. Leur intérêt ludique, leur efficacité, et/ou la popularité, la couverture médiatique du projet dont ils sont l'instrument – les quatre, dans l'idéal – doivent non seulement attirer, mais fidéliser un nombre suffisant de joueurs pour que la ressource produite soit à la fois importante et de qualité. Et ceci en visant le public le plus large possible tout en tenant compte de sa diversité, puisque pour certains, l'aspect ludique va largement primer, tandis que d'autres seront davantage séduits par l'idée de participer à quelque chose d'utile : la *science citoyenne* a le vent en poupe, et beaucoup de gens sont prêts à s'impliquer dans ce type de projet, et ce même si ils n'en maîtrisent pas clairement les tenants et les aboutissants, mais simplement parce que c'est valorisant.

Si on se concentre sur les aspects jouabilité/efficacité, notons que pour un GWAP, aux aspects de conception traditionnelle d'un jeu sur Internet s'ajoute la difficulté d'avoir une boucle ludique covariante avec la boucle d'acquisition de données. Cela signifie qu'il faut s'efforcer de rendre inter-dépendants, à travers le mécanisme du jeu, l'intérêt du joueur et la qualité des données qu'il produit. Autrement dit, un joueur qui joue bien (donc qui produit des données de qualité) doit être récompensé par le jeu afin d'encourager cette façon de jouer, alors que l'absence de récompense, voire la pénalisation pour le joueur qui joue mal doivent être couplées avec le mécanisme qui évite de collecter de mauvaises données.

En ce qui concerne l'acquisition de données lexico-sémantiques, le recours aux GWAP, relativement récent puisque lié à l'émergence d'Internet, ne vient pas en opposition, mais en complément des méthodes traditionnelles, essentiellement manuelles, qui existent depuis plusieurs siècles. Ainsi, le réseau lexical obtenu grâce au projet *JeuxDeMots* n'a pas pour vocation de supplanter des ressources comme WordNet ou les dictionnaires d'usage, mais plus modestement d'être un complément à ces outils de référence, et de démontrer la réelle efficacité d'approches fondées sur de la peuplologie ludique. Certains autres jeux, principalement ceux de type macrotâche, ne constituent pas, par contre, un complément aux méthodes traditionnelles, mais bien une alternative susceptible de permettre l'acquisition de données ou la résolution de problèmes que l'on ne sait ou ne peut envisager *via* les méthodes traditionnelles.

Dans le cadre du projet *JeuxDeMots*, nous avons observé que la collecte d'informations, qu'elles relèvent du lexique ou du sens commun, se faisait aisément *via* des jeux, et avec une qualité très satisfaisante. Les auteurs de GWAP qui ont publié des estimations de la qualité des ressources obtenues sont très majoritairement satisfaits. Nous avons cependant constaté que plus l'information demande une expertise poussée, plus les résultats sont aléatoires. Par exemple, en ce qui concerne la constitution d'un réseau lexical, peu de données exigent une compétence linguistique dépassant l'intuition du locuteur ; par contre, des tâches plus complexes, comme la constitution d'une banque d'arbres d'analyses syntaxiques, ou des alignements de traduction s'adressent à un public beaucoup plus restreint dans la mesure où elles nécessitent des connaissances bien plus spécialisées. Le constat est le même pour deux GWAP médicaux, *Dizeez* et *The Cure*, que nous décrivons au chapitre 2. Notons enfin qu'un certain nombre de ressources ainsi acquises par GWAP peuvent avantageusement être évaluées *via*... un autre GWAP, qui doit bien entendu présenter des modalités orthogonales à celles d'acquisition. Un des avantages majeurs de ce genre d'approche pour l'évaluation est que l'échantillonnage est effectué implicitement par les joueurs eux-mêmes.

Ainsi, *la boucle est bouclée* : ce sont les joueurs qui, par une activité ludique, ont permis d'acquérir une ressource, ce sont également les joueurs, toujours *en s'amusant*, qui valident la ressource acquise. Nul ne doute que Galilée, voyant les résultats obtenus par cette *boucle vertueuse*, réaffirmerait cette fois-ci haut et fort « E pur si muove ! ».

Version auteurs, dec. 2014

Bibliographie

- [von Ahn et Dabbish, 2008] von Ahn L., Dabbish L. « Designing Games With a Purpose », *Communication of the ACM*, p. 58-67, 51(8), 2008.
- [von Ahn et al., 2006] von Ahn L., Liu R., Blum M. « Peekaboom : a game for locating objects in images », *Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems*, p. 55-64, Montréal, Canada, 2006.
- [von Ahn, 2006] von Ahn L. « Games with a purpose », *Computer*, p. 92-94, 39(6), 2006.
- [von Ahn et al., 2006] von Ahn L., Kedia M., Blum M. « Verbosity : a game for collecting common-sense facts », *Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems*, volume 1 of Games, p. 75-78, Montréal, Canada, 2006.
- [von Ahn et Dabbish, 2004] von Ahn L., Dabbish L., « Labeling images with a computer game », *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, Etats-Unis, ACM Press, p. 319-326, 2004.
- [Alonso et Mizzaro, 2009] Alonso O., Mizzaro S. « Can we get rid of trec assessors ? using mechanical turk for relevance assessment », *SIGIR '09 : Workshop on The Future of IR Evaluation*, Boston, Etats-Unis, 2009.
- [Alonso et al., 2008] Alonso O., Rose D. E., Stewart B. « Crowdsourcing for relevance evaluation », *SIGIR Forum*, p. 9-15, 42(2), 2008.
- [Artstein et Poesio, 2008] Artstein R., Poesio M. « Inter-coder agreement for computational linguistics », *Computational Linguistics*, p. 555-596, 34(4), 2008.
- [Attardi et the Galoap Team, 2010] Attardi G. et the Galoap Team « Phratris - A Phrase Annotation Game », *Demo presented at Insemtives Game Idea Challenge*, 2010.
- [Callison-Burch, 2009] Callison-Burch C. « Fast, cheap, and creative : Evaluating translation quality using amazon's mechanical turk », *Proceedings of the 2009 Conference on Empirical Methods in*

- Natural Language Processing, EMNLP*, Association for Computational Linguistics. p. 286-295, volume 1, Singapour, 2009.
- [Chamberlain *et al.*, 2013] Chamberlain J., Fort K., Kruschwitz U., Lafourcade M., Poesio M. « Using games to create language resources : successes and limitations of the approach », *Theory and Applications of Natural Language Processing*, dir. I. Gurevych et J. Kim, Springer, 2013.
- [Chamberlain *et al.*, 2012] Chamberlain J., Kruschwitz U., Poesio M. « Motivations for participation in socially networked collective intelligence systems », *Proceedings of CI2012*, Cambridge, Etats-Unis, 2012.
- [Chamberlain *et al.*, 2009] Chamberlain J., Poesio M., Kruschwitz U. « A new life for a dead parrot : Incentive structures in the Phrase Detectives game », *Proceedings of the WWW 2009 Workshop on Web Incentives (WebCentives'09)*, Madrid, Espagne, 2009.
- [Chamberlain *et al.*, 2009] Chamberlain J., Poesio M., Kruschwitz U. « A demonstration of human computation using the Phrase Detectives annotation game », *HCOMP '09 : Proceedings of the ACM SIGKDD Workshop on Human Computation*, p. 23-24, Paris, France, 2009.
- [Chamberlain *et al.*, 2008] Chamberlain J., Poesio M., Kruschwitz U. « Phrase Detectives : A Web-based Collaborative Annotation Game », *Proceedings of the International Conference on Semantic Systems (I-Semantics'08)*, Graz, Autriche, 2008.
- [Chklovski et Gil, 2005] Chklovski T., Gil Y. « Improving the design of intelligent acquisition interfaces for collecting world knowledge from web contributors », *Proceedings of K-CAP '05, ACM*, p. 35-42, New York, Etats-Unis, Association for Computing Machinery, 2005.
- [Chklovski, 2005] Chklovski T. « Collecting paraphrase corpora from volunteer contributors », *Proceedings of K-CAP '05, ACM*, p. 115-120, New York, Etats-Unis, Association for Computing Machinery, 2005.
- [Cooper *et al.*, 2010] Cooper S., Khatib F., Treuille A., Barbero J., Lee J., Beenen M., Leaver-Fay A., Baker D., Popovic Z. and Foldit players « Predicting protein structures with a multiplayer online game », *Nature* p. 756-760, 466, 2010.
- [Dong et Dong, 2006] Dong Z., Dong Q. « HowNet and the Computation of Meaning », *World-Scientific*, Londres, Angleterre, 2006.
- [Fabre et Bourigault, 2006] Fabre C., Bourigault D. « Extraction de relations sémantiques entre noms et verbes au-delà des liens morphologiques », *Proceedings of TALN'06*, Louvain, Belgique, p. 121-126, 2006.
- [Fellbaum, 1998] Fellbaum C. « Wordnet an electronic lexical database », Cambridge, Etats-Unis, MIT Press, 1998.

- [Feng *et al.*, 2009] Feng D., Besana S., Zajac R. « Acquiring high quality non-expert knowledge from on-demand workforce », *Proceedings of the 2009 Workshop on The People's Web Meets NLP : Collaboratively Constructed Semantic Resources, People's Web '09*, pages 51-56, Morristown, Etats-Unis, Association for Computational Linguistics, 2009.
- [Fisher, 2009] Fisher L. « The Perfect Swarm : the Science of Complexity in Everyday Life », *Basic Books*, p. 260, 2009.
- [Fort *et al.*, 2014] Fort K., Guillaume B., Chastant H. « Creating Zombilingo, a game with a purpose for dependency syntax annotation », *Proceedings of the First International Workshop on Gamification for Information Retrieval, GamifIR'14*, Amsterdam, Pays-Bas, 2014.
- [Fort *et al.*, 2011] Fort K., Adda G., Cohen K. B. « Amazon mechanical turk : Gold mine or coal mine ? », *Computational Linguistics*, 37(2), 2011.
- [Gala et Zock, 2014] Gala N., Zock M. (Eds) : *Ressources Lexicales - Contenu, construction, utilisation, évaluation*, John Benjamins Publishing Company, 2013.
- [Good *et al.*, 2014] Good B. M., Loguercio S., Griffith O. L., Nanis M., Wu C., Su A. I., « The Cure : Design and Evaluation of a Crowdsourcing Game for Gene Selection for Breast Cancer Survival Prediction », *JMIR Serious Games*, 2014, 2(2) : e7, 2014.
- [Good et Su, 2013] Good B. M., Su A. I. « Crowdsourcing for Bioinformatics », *Bioinformatics* p. 1925-1933, 29 (16) 2013.
- [Green *et al.*, 2010] Green N., Breimyer P., Kumar V., Samatova N.F. « PackPlay : Mining semantic data in collaborative games », *Proceedings of the 4th Linguistics Annotation Workshop, LAW IV'10*, Uppsala, Suède, 2010.
- [Hand, 2010] Hand E. « Citizen science : People power », *Nature*, p. 685-687, 466, 4 août 2010.
- [Hladka *et al.*, 2011] Hladka B., Mirovsky J., Kohout J. « An attractive game with the document : (im)possible ? », *The Prague Bulletin of Mathematical*, p. 5-26, (96) 2011.
- [Hladka *et al.*, 2009] Hladka B., Mirovsky J., Schlesinger P. « Play the language : play coreference », *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, p. 209-212, Singapour, Association for Computational Linguistics, 2009.
- [Johnson *et al.*, 1998] Johnson N. L., Rasmussen S., Joslyn C., Rocha L., Smith S., Kantor M. « Symbiotic Intelligence : Self-Organizing Knowledge on Distributed Networks Driven by Human Interaction », *Proceedings of the Sixth International Conference on Artificial Life*, MIT Press, 1998.
- [Joubert et Lafourcade, 2012] Joubert A., Lafourcade M. « A new dynamic approach for lexical networks evaluation », *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turquie, 23-25 mai 2012.

- [Joubert *et al.*, 2011] Joubert A., Lafourcade M., Schwab D., Zock M. « Évaluation et consolidation d'un réseau lexical grâce à un assistant ludique pour le " mot sur le bout de la langue " », *Proceedings of TALN'11*, Montpellier, France, p. 295-306, 2011.
- [Jurafsky et Martin, 2008] Jurafsky D., Martin J. H. « Speech and Language Processing », 2nd edition. Prentice-Hall, 2008.
- [Jurgens et Navigli, 2014] Jurgens D., Navigli R. « It's All Fun and Games until Someone Annotates : Video Games with a Purpose for Linguistic Annotation », *Transactions of the Association for Computational Linguistics (TACL)*, p. 449-464, 2, 2014.
- [Kanefsky *et al.*, 2001] Kanefsky B., Barlow N., Gulick V. « Can distributed volunteers accomplish massive data analysis tasks ? », *Lunar and Planetary Science*, XXXII, 32, 2001.
- [Kawrykow *et al.*, 2012] Kawrykow A., Roumanis G., Kam A., Kwak D., Leung C., Wu C., Zaurour E., Phylo Players, Sarmenta L., Blanchette M., Waldspühl J. « Phylo : A Citizen Science Approach for Improving Multiple Sequence Alignment », *PLoS ONE*, 7(3) : e31362, 2012.
- [Kazai, 2011] Kazai G. « In Search of Quality in Crowdsourcing for Search Engine Evaluation », *Proceedings of the 33rd European Conference on Information Retrieval (ECIR'11)*, Lecture Notes in Computer Science, Springer, p. 165-176, volume 6611, 2011.
- [Kazai *et al.*, 2009] Kazai G., Milic-Frayling N., Costello J. « Towards methods for the collective gathering and quality control of relevance assessments », *Proceedings of the 32nd international ACM SIGIR conference on Research and development in Information Retrieval, SIGIR '09*, p. 452-459, New York, Etats-Unis, ACM, 2009.
- [Khatib *et al.*, 2011] Khatib F., Cooper S., Tyka M.D., Xu K., Makedon I., Popovic Z., Baker D. and Foldit Players . « Algorithm discovery by protein folding game players », *PNAS*, p. 18949-18953, 108 (47), 2011.
- [Kim *et al.*, 2014] Kim J.S., Greene M.J., Zlateski A., Lee K., Richardson M., Turage S.C., Purcaro M., Balkam M., Robinson A., Behabadi B.F., Campos M., Denk W., Seung H.S., and the EyeWriters « Space-time wiring specificity supports direction selectivity in the retina », *Nature* p. 331-336, 509, 2014.
- [Koller *et al.*, 2010] Koller A., Striegnitz K., Gargett A., Byron D., Cassell J., Dale R., Moore J., Oberlander J. « Report on the second nlg challenge on generating instructions in virtual environments (give-2) », *Proceedings of the 6th INLG*, Dublin, Irlande, 2010.
- [Koster, 2005] Koster R. « A Theory of Fun for Game Design », *Paraglyph*, 2005.
- [Kuo *et al.*, 2009] Kuo Y.L., Lee J.C., Chiang K.Y., Wang R., Shen E., Chan C.W., Yung-jen Hsu J. « Community-based game design : experiments on social games for commonsense data collection », *Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP'09*, p. 15-22, New York, Etats-Unis, 2009.

- [Kwak *et al.*, 2013] Kwak D., Kam A., Becerra D., Zhou Q., Hops A., Zarour E., Kam A., Sarmenta L., Blanchette M., Waldspühl J. « Open-Phylo : a customizable crowd-computing platform for multiple sequence alignment », *Genome Biology* 14 : R116, 2013.
- [Lafourcade *et al.*, 2014] Lafourcade M., Le Brun N., Zampa V. « Crowdsourcing word-color associations », *Proceedings of the 19th International Conference on Application of Natural Language to Information Systems (NLDB 2014)*, Montpellier, France, 6 p., 18-20 juin 2014.
- [Lafourcade et Fort, 2014] Lafourcade M., Fort K. « Propa-L : a semantic filtering service from a lexical network created using Games With A Purpose », *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*, Reykjavik, Islande, p. 1676-1681, 26-31 mai 2014.
- [Lafourcade et Joubert, 2010] Lafourcade M., Joubert A. « Computing Trees of Named Word Usages from a Crowdsourced Lexical Network », *Investigationes Linguisticae*, p. 39-56, vol. XXI, 2010.
- [Lafourcade et Zampa, 2009] Lafourcade M., Zampa V. « PtiClic : a game for vocabulary assessment combining JeuxDeMots and LSA », *Proceedings of CICLing (Conference on Intelligent text processing and Computational Linguistics)*, Mexico, 2009.
- [Lafourcade, 2007] Lafourcade M. « Making People Play for Lexical Acquisition », *Proceedings of SNLP 2007, 7th Symposium on Natural Language Processing*, Pattaya, Thaïlande, 8 p., 2007.
- [Lee *et al.*, 2014] Lee J., Kladwang W., Lee M., Cantu D., Azizyan M., Kim H., Limpaecher A., Yoon S., Treuille A., Das R. and EteRNA Participants « RNA design rules from a massive open laboratory », *PNAS*, p. 2122-2127, 111 (6), 2014.
- [Lenat, 1995] Lenat D. « CYC : A large-scale investment in knowledge infrastructure », *Communications of the ACM*, p. 33-38, 38(11), 1995.
- [Lieberman *et al.*, 2007] Lieberman H., Smith D., Teeters A. « Common Consensus : A Web-based Game for Collecting Commonsense Goals », *Workshop on Common Sense for Intelligent Interfaces, ACM International Conference on Intelligent User Interfaces (IUI-07)*, Honolulu, Hawaii, Etats-Unis, 2007.
- [Luengo-Oroz *et al.*, 2012] Luengo-Oroz M.A., Arranz A., Frean J. « Crowdsourcing Malaria Parasite Quantification : An Online Game for Analyzing Images of Infected Thick Blood Smears », *JMIR*, 14 (6) : e167, 2012.
- [Malone *et al.*, 2009] Malone T.W., Laubacher R., Dellarocas C. « Harnessing crowds : Mapping the genome of collective intelligence », *MIT Sloan Research Paper No. 4732-09*, p. 327-335, 2009.

- [Marcus *et al.*, 1993] Marcus M., Santorini B., Marcinkiewicz M.A. « Building a large annotated corpus of english : the Penn treebank », *Computational Linguistics*, p. 313-330, 19(2), 1993.
- [Marshall, 2012] Marshall J. « Victory for crowdsourced biomolecule design », *Nature News*, 1038/nature 9872, 2012.
- [Mason et Watts, 2010] Mason W., Watts D.J. « Financial incentives and the "performance of crowds" », *SIGKDD Explor. Newsl.*, p. 100-108, 11, 2010.
- [Mavandadi *et al.*, 2012] Mavandadi S., Dimitrov S., Feng S., Yu F., Sikora U., Yaglidere O., Padmanabhan S., Nielsen K., Ocan A. « Distributed Medical Image Analysis and Diagnosis through Crowd-Sourced Games : a Malaria Case Study », *PLoS ONE*, 7(5) : e37245, 2012.
- [Michael et Chen, 2006] Michael D.R., Chen S. « Serious Games : Games that Educate, Train and Inform », *Thomson Course Technology*, 287 p., 2006.
- [Mihalcea et Chklovski, 2003] Mihalcea R., Chklovski T. « Open MindWord Expert : Creating large annotated data collections with web users help », *Proceedings of the EACL 2003, Workshop on Linguistically Annotated Corpora (LINC 2003)*, 2003.
- [Miller, 1995] Miller G. « WordNet : A Lexical Database for English », *Communications of the ACM*, p. 39-41, vol. 38, No. 11, 1995.
- [Miller *et al.*, 1990] Miller G.A., Beckwith R., Fellbaum C., Gross D., Miller K.J. « Introduction to WordNet : an on-line lexical database », *International Journal of Lexicography*, p. 235-244, 3 (4), 1990.
- [Mrozinski *et al.*, 2008] Mrozinski J., Whittaker E., Furu S. « Collecting a whyquestion corpus for development and evaluation of an automatic QA-system », *Proceedings of ACL-08 : HLT*, p. 443-451, Columbus, Etats-Unis, Association for Computational Linguistics, 2008.
- [Navigli et Ponzetto, 2010] Navigli R., Ponzetto S. « BabelNet : Building a very large multilingual semantic network », *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Suède, p. 216-225, 2010.
- [Ploux et Victorri, 1998] Ploux S., Victorri B. « Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes », *Traitement Automatique des Langues*, p. 161-182, vol. 39/1, 1998.
- [Poesio *et al.*, 2006] Poesio M., Sturt P., Arstein R., Filik R. « Underspecification and anaphora : Theoretical issues and preliminary evidence », *Discourse Processes*, p.157-175, 42(2), 2006.
- [Poesio, 2004] Poesio M. « Discourse annotation and semantic annotation in the GNOME corpus », *2004 ACL Workshop on Discourse Annotation*, Stroudsburg, Etats-Unis, 2004.
- [Pustejovsky, 1993] Pustejovsky J. « The generative lexicon », *Computational Linguistics*, p. 409-441, 1993.

- [Quinn et Bederson, 2011] Quinn A., Bederson B. « Human computation : a survey and taxonomy of a growing field », *Proceedings of ACM CHI 2011 Conference on Human Factors in Computing Systems*, Vancouver, Canada, 2011.
- [Rafelsberger et Scharl, 2009] Rafelsberger W., Scharl A. « Games with a purpose for social networking platforms », *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, p. 193-198, Turin, Italie, ACM, 2009.
- [Sagot *et al.*, 2011] Sagot B., Fort K., Adda G., Mariani J., Lang B. « Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé », *Proceedings of Traitement Automatique des Langues Naturelles (TALN 2011)*, Montpellier, France, 27 juin-1er juillet 2011.
- [Sagot et Fiser, 2008] Sagot B., Fiser D. « Construction d'un wordnet libre du français à partir de ressources multilingues », *Proceedings of TALN 2008*, Avignon, France, 2008.
- [Schrope, 2013] Schrope M. « Solving thought problems with games », *PNAS* p. 7104-7106, 110 (18) 2013.
- [Siorpaes et Hepp, 2008] Siorpaes K., Hepp M. « Games with a Purpose for the Semantic Web », *IEEE Intelligent Systems*, p. 50-60, vol. 23, No. 3, 2008.
- [Smadja, 2009] Smadja F. « Mixing financial , social and fun incentives for social voting », *World Wide Web Internet And Web Information Systems*, avril 2009.
- [Snow *et al.*, 2008] Snow R., O'Connor B., Jurafsky D., Ng A.Y. « Cheap and fast—but is it good ? : evaluating non-expert annotations for natural language tasks », *EMNLP'08 : Proceedings of the Conference on Empirical Methods in Natural Language Processing, ACL*, p. 254-263, Morristown, Etats-Unis, Association for Computational Linguistics, 2008.
- [Steinmayr *et al.*, 2011] Steinmayr B., Wieser C., Kneißl F., Bry F. « Karido : A GWAP for Telling Artworks Apart », *Proceedings of the 16th International Conference on Computer Games*, 27-30 juillet 2011, p. 193-200, 2011.
- [Störkle, 2012] Störkle F. « Combino - A GWAP for Generating Combined Tags », *Rapport de stage de MASTER, Institut d'Informatique - LMU - Munich*, 53 p., 2012.
- [Strauss, 2012] Strauss S. « Gamers outdo computers at matching up disease genes - Computer game crowdsources DNA sequence alignments across different species », *Nature News* (2012) 1038/nature.10203, 2012.
- [Tchernia, 2014] Tchernia R. « Science participative, les joueurs aux manettes », *Le Monde*, 5 mars 2014.
- [Thaler *et al.*, 2011] Thaler S., Siorpaes K., Simperl E., Hofer Ch. « A Survey on Games for Knowledge Acquisition », *STI Technical Report*, 2011.

- [Vannella *et al.*, 2014] Vannella D., Jurgens D., Scarfini D., Toscani D., Navigli R. « Validating and extending semantic knowledge bases using video games with a purpose », *Proceedings of the Annual Meeting of the Association for Computational Linguistics, ACL'14*, 2014.
- [Venhuizen *et al.*, 2013] Venhuizen N., Basile V., Evang K., Bos J. « Gamification for Word Sense Labeling », *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Short Papers*, p. 397-403, Potsdam, Allemagne, 2013.
- [Véronis et Ide, 1990] Véronis J., Ide N. « Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries », *Proceedings of 13th International Conference on Computational Linguistics (COLING'90)*, p. 389-394 vol. 2, Helsinki, Finlande, 1990.
- [Vickrey *et al.*, 2008] Vickrey D., Bronzan A., Choi W., Kumar A., Turner-Maier J., Wang A., Koller D. « Online Word Games for Semantic Data Collection », *Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP'08*, p. 533-542, Honolulu, Etats-Unis, octobre 2008.
- [Vossen, 1998] Vossen P. « EuroWordNet : A Multi-lingual Database with Lexical Semantic Networks », *Kluwer*, Dordrecht, Pays-Bas, 1998.
- [Wieser *et al.*, 2013] Wieser C., Bry F., Bérard A., Lagrange R. « ARTigo : Building an Artwork Search Engine With Games and Higher-Order Latent Semantic Analysis », *First AAAI Conference on Human Computation and Crowdsourcing*, Palm Springs, Etats-Unis, 2013.
- [Zampa et Lafourcade, 2010] Zampa V., Lafourcade M. « PtiClic et PtiClic-kids : jeux avec les mots permettant une double acquisition », *Proceedings of TICE2010, 7e colloque TICE*, Nancy, France, 2010.
- [Zesch et Gurevych, 2009] Zesch T., Gurevych I. « Wisdom of crowds versus wisdom of linguists measuring the semantic relatedness of words », *Natural Language Engineering*, Cambridge University Press, p. 25-59, 2009.

Index

1001 Paraphrases, 73

Akinator, 92

Amazon Mechanical Turk, 2

AMT, 2

annotation, 59, 62

Apetopia, 81

apprentissage des langues, 83

ARN, 11

ARTigo, 86

Askit, 106

Asku, 116

association entre termes, 69, 99, 103

Barbora Hladka, 62

Be a martian, 90

Beat the Bots, 79

Bruno Guillaume, 67

camouflage, 30

cancer du sein, 39

captcha, 2

Categozilla, 73

Categorilla, 73

Center for Game Science, 8

chercher un œuf, 30

Citizen sort, 27

ColorIt, 111

Combino, 88

Common Consensus, 75

coréférences, 62

couleurs, 81, 111

désambiguïsation lexicale, 71

Daniele Vannella, 69

David Vickrey, 73, 74

description d'œuvres d'art, 86, 88, 89

Diko, 116

Dizeez, 38

Duolingo, 83

Egglab game, 32

Emot, 112

émotions, 112

Entity Discovery, 74

environnement contributif, 116

ESP Game, 2

EteRNA, 11

Eyewire, 23

FACTory Game, 76

Find the link, 77

Foldit, 8
 Forgotten Island, 30
 Fraxinus, 21
 FreeAssociation, 74

 GalaxyZoo, 4
 Giuseppe Attardi, 75
 Groningen Meaning Bank, 71
 GuessWhat ?, 77

 Happy match, 28
 Henry Lieberman, 75

 Infection, 69

 jeu sérieux, 1
 jeux à sélections multiples, 111
 jeux à votes, 105
 jeux biologiques, 7
 Citizen sort, 27
 EteRNA, 11
 Eyewire, 23
 Foldit, 8
 Forgotten Island, 30
 Fraxinus, 21
 Happy match, 28
 Nanocrafter, 14
 Phylo, 17
 projet Nightjar
 Egglab game, 32
 Nest game, 32
 jeux d'associations, 99
 jeux d'attribution, 103
 jeux d'ordonancement, 114
 jeux de devinettes, 92, 104
 jeux inclassables, 79
 Akinator, 92

 Apetopia, 81
 ARTigo, 86
 Be a martian, 90
 Beat the Bots, 79
 Combino, 88
 Duolingo, 83
 Karido, 89
 Quantum Moves, 82
 jeux médicaux, 35
 Dizeez, 38
 Malaria Spot Game, 43
 Malaria Training Game, 41
 Nanodoc, 35
 Play to Cure : Genes in Space, 48
 The Cure, 39
 Worm Watch Lab, 46
 jeux pour le TAL, 51
 1001 Paraphrases, 73
 Categodzilla, 73
 Categorilla, 73
 Common Consensus, 75
 Entity Discovery, 74
 FACTory Game, 76
 Find the link, 77
 FreeAssociation, 74
 GuessWhat ?, 77
 Infection, 69
 JeuxDeMots, 66
 OnToGalaxy, 76
 Open Mind Word Expert, 72
 PhraseDetectives, 59
 PhraTris, 75
 PlayCoref, 62
 Train Robots, 76
 Verbosity, 64
 Wordrobe, 71

Zombilingo, 67
 jeux sérieux, 1
 JeuxDeMots, 66
 Jiri Mirovsky, 62
 Johan Bos, 71
 Jon Chamberlain, 59

 Karën Fort, 67
 Karido, 89

 lien anaphorique, 59
 LikeIt, 108
 Luis Von Ahn, 2, 84

 macrotâches, 4
 malaria, 41
 Malaria Spot Game, 43
 Malaria Training Game, 41
 Manuel Blum, 64
 Markus Krause, 76
 Mars, 90
 Massimo Poesio, 59
 microtâches, 3
 Mihir Kedia, 64
 molécules d'ARN, 11
 mot sur le bout de la langue, 104

 Nanocrafter, 14
 Nanodoc, 35
 Nathan Green, 74
 Nest game, 32
 Nightjar, 30

 OCR, 2
 OnToGalaxy, 76
 Open Mind Word Expert, 72
 opinion, 108

 paludisme, 41
 personnage, 92
 PhraseDetectives, 59
 PhraTris, 75
 Phylo, 17
 physique quantique, 82
 planète Mars, 90
 Play to Cure : Genes in Space, 48
 PlayCoref, 62
 polarité, 108
 politique, 113
 PolitIt, 113
 projet JeuxDeMots, 97, 99
 Askit, 106
 Asku, 116
 ColorIt, 111
 Diko, 116
 Emot, 112
 environnement contributif, 116
 jeux à votes, 105
 JeuxDeMots, 99
 LikeIt, 108
 PolitIt, 113
 PtiClic, 103
 SexIt, 110
 Tierxical, 114
 Totaki, 104
 projet Nightjar, 30
 Egglab game, 32
 Nest game, 32
 pseudo-jeu d'apprentissage, 116
 PtiClic, 103

 Quantum Moves, 82

 réseau lexical, 97
 Rada Mihalcea, 73

ReCAPTCHA, 2
reconnaissance optique de caractères, 2
Roberto Navigli, 69

Sangeeta Bhatia, 35
science citoyenne, 3
sens commun, 64
sentiments, 112
serious game, 1
sexe, 110
SexIt, 110
shoot 'em up, 69

Tetrisi, 75
The Cure, 39
tiercé, 114
Tierxical, 114
Timothy Chklovski, 73
Totaki, 104
traduction, 83
Train Robots, 76

Udo Kruschwitz, 59

Verbosity, 64
VouchSafe, 79

Wordrobe, 71
Worm Watch Lab, 46

Zombilingo, 67
Zooniverse, 4, 46
Zoran Popovic, 14

Jeux et intelligence collective

résolution de problèmes et acquisition de données sur le Web

Résumé

Les jeux avec but (GWAP en anglais) sont des jeux à travers lesquels l'activité ludique permet de collecter des données ou de résoudre des problèmes trop complexes, ou trop coûteux en termes de moyens humains et matériels pour être résolus par des machines. Ethiquement plus acceptables que bon nombre d'approches par externalisation ouverte (crowdsourcing), ces jeux, qui représentent un type particulier de jeux sérieux, n'en restent pas moins délicats à concevoir, car ils doivent non seulement être attrayants, mais efficaces, c'est-à-dire permettre de récolter des données de bonne qualité, ou encore de progresser significativement dans la résolution d'un problème. Dans cet ouvrage, nous montrons que le concept de GWAP, actuellement en pleine expansion, s'illustre par une grande diversité de formes et d'applications, et se décline dans de multiples domaines, en particulier en biologie, en médecine, et dans le domaine du Traitement Automatique du Langage. Spécialisés dans ce dernier domaine, et nous-mêmes à l'origine, depuis 2007, d'un ensemble de GWAP visant à construire un réseau lexical, (projet JeuxDeMots, auquel nous consacrons un chapitre), nous tentons de définir, sur la base de notre enquête sur les GWAP existant, et de notre propre expérience, les critères et conditions qu'il est important de privilégier dans la conception d'un GWAP afin d'en faire un outil d'acquisition de données efficace.

Mots-clés

GWAP, jeux avec but, acquisition de données, résolution de problème, crowdsourcing, externalisation ouverte, TAL, traitement automatique du langage, JeuxDeMots.

Les auteurs

Mathieu Lafourcade est enseignant-chercheur en informatique et Traitement des Langues, et est l'inventeur du projet JeuxDeMots. Ses centres d'intérêt vont de l'analyse sémantique de textes par des colonies de fourmis virtuelles, à l'inférence et le raisonnement sur des réseaux lexicaux.

Nathalie Le Brun, docteur en Biologie, ingénieur CNRS dans une autre vie, est aujourd'hui graphiste, experte en jeux sérieux et GWAP, et participe activement au projet JeuxDeMots. Dans une vie future, elle sera peut être linguiste et lexicologue.

Alain Joubert est maître de conférences et effectue ses activités de recherche au LIRMM. Docteur en Astrophysique, ses principaux thèmes de recherche actuels concernent le traitement automatique du langage. Il collabore depuis sept ans au projet JeuxDeMots.