

# John Benjamins Publishing Company



This is a contribution from *Ressources Lexicales. Contenu, construction, utilisation, évaluation.*

Edited by Núria Gala and Michael Zock.

© 2013. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: [www.copyright.com](http://www.copyright.com)).

Please contact [rights@benjamins.nl](mailto:rights@benjamins.nl) or consult our website: [www.benjamins.com](http://www.benjamins.com)

Tables of Contents, abstracts and guidelines are available at [www.benjamins.com](http://www.benjamins.com)

## Bénéfices et limites de l'acquisition lexicale dans l'expérience JeuxDeMots

Mathieu Lafourcade et Alain Joubert  
LIRMM – Université Montpellier 2

### Introduction

L'acquisition de ressources lexicales est un problème récurrent en TALN. Celle-ci peut se faire automatiquement, généralement à partir de textes, mais dans ce cas elle peut être biaisée par le corpus choisi (généraliste ou spécialisé, journalistique ou dictionnaire, ...), aussi bien que par la méthode d'acquisition qui peut favoriser l'émergence de certaines informations au détriment d'autres. De plus, l'extraction automatique de relations typées entre termes à l'aide de patrons syntaxiques se heurte à leur ambiguïté intrinsèque et résulte généralement en un bruit important nécessitant une correction au cas par cas (voir par exemple (Fabre & Bourigault 1998) pour l'extraction de relations entre noms et verbes). L'acquisition de ressources peut également être effectuée selon la tradition lexicographique, à savoir de façon manuelle. Dans ce cas, plusieurs méthodes sont possibles, les ressources pouvant être créées : (a) par des experts : c'est alors un processus long et coûteux, souvent lourd à faire évoluer ; (b) par des approches contributives rémunérées ou non sollicitant des participants non nécessairement linguistes de formation *mais attirés par les choses de la langue* ; (c) enfin par une stratégie plus récente consistant à mettre en œuvre des jeux accessibles sur Internet. Ces jeux permettent d'acquérir des ressources par consensus populaire entre locuteurs natifs. C'est un processus rapide, peu coûteux, très dynamique, mais cependant la question de la qualité de la ressource peut légitimement être posée.

Toutefois, selon (Fisher 2009), il s'avère, en général, qu'une estimation donnée par un groupe d'individus non experts qui n'interagissent pas entre eux, est souvent meilleure que les choix proposés par un groupe d'experts élaborant collectivement leurs réponses. Par ailleurs, un avis moyen obtenu par un grand nombre d'individus a plus de chance de se rapprocher de l'optimum (ou d'être

représentatif) que par un groupe de plus petite taille. On notera dans cet esprit les travaux de (Zesch & Gurevych 2009) concernant la similarité sémantique entre termes. A nouveau, cette dernière observation n'est plus strictement valide si la réponse est négociée au sein du groupe, plutôt que calculée par moyenne des réponses individuelles. Les jeux en ligne visant à construire des ressources exploitent implicitement ces considérations, les joueurs menant une activité isolée. Cette activité répétée un grand nombre de fois avec un grand nombre de personnes permet de faire émerger des données de qualité. Par ailleurs, l'approche ludique est une condition sans doute essentielle à l'obtention d'un grand nombre de participants motivés. Toutefois, il reste à établir fermement la validité de ces hypothèses concernant la collecte de ressources lexicales.

## 1. Contexte

### 1.1 Acquisition lexicale par peuplonomie

Les jeux visant à construire des ressources restent rares même si ce domaine est en expansion. Dans la littérature, il est souvent mentionné de *Human Computation* lorsque l'activité (plus ou moins ludique) consiste à substituer des utilisateurs à la machine, pour des tâches en général difficilement automatisables de façon satisfaisante. Si le nombre d'utilisateurs est important, il s'agit alors de *Crowdsourcing* (*Peuplonomie*), que l'approche soit contributive ou ludique. Il est à noter que les approches contributives rémunérées à la tâche posent un certain nombre de questions quant à leur efficacité et l'éthique, notamment dans le cadre d'Amazon Mechanical Turk (Fort et al. 2011, Sagot et al. 2011). Concernant les données de nature lexicale, sémantiques ou de connaissance du monde, les exemples réels d'approches ludiques se comptent toutefois sur les doigts d'une main.

ESP (nom choisi en référence à Extrasensory Perception) (von Ahn & Dabbish 2004) est un jeu où deux participants doivent décrire une image à l'aide de mots-clés. Les joueurs ne sont pas en mesure de communiquer. Des points sont gagnés lorsque les termes des deux joueurs correspondent. L'idée est que les termes fournis par les joueurs constituent des éléments d'indexation pertinents pour décrire l'image. Ne retenir que les mots clés communs fournit la base d'un filtrage minimal permettant d'éviter la majeure partie du bruit. A notre connaissance, ESP a été le premier jeu en ligne dont l'objet était de récupérer de telles données via l'activité des joueurs. D'autres jeux ont vu le jour, notamment concernant la localisation d'objets dans une image (von Ahn et al. 2006a), la collecte d'informations de sens commun (Lieberman et al. 2007, von Ahn et al. 2006b, Kuo et al. 2009), ainsi que (Siorpaes & Hepp 2008) pour la construction/validation d'ontologies.

*Phrase Detective* (Chamberlain et al. 2009) est un jeu coopératif en ligne permettant la constitution d'un corpus annoté concernant l'anaphore. L'objet d'une séquence de jeu est, à partir d'un texte et d'éléments de ce texte mis en évidence, de déterminer la bonne référence dans ce même texte. Le jeu est proposé au joueur alternativement en identification ou en validation de propositions faites par d'autres joueurs. Le joueur gagne autant de points que le nombre de personnes ayant fourni une réponse identique à la sienne. Il s'agit en définitive d'une forme de sondage à questions semi-fermées ludicisé.

D'une façon générale, les données produites par les jeux à notre connaissance ne sont pas librement disponibles. Il est possible de tenter d'analyser les mécanismes des approches proposées, mais il reste difficile voire impossible de confronter ces approches aux données effectivement récoltées ou construites.

## 1.2 Les réseaux lexicaux

Les réseaux lexicaux, qu'ils soient généraux ou spécialisés, constituent, en particulier depuis l'émergence de WordNet (Miller et al. 1990, Vossen 1998 pour EuroWordNet) une ressource lexicale considérée comme particulièrement précieuse aux applications de l'informatique linguistique, dont notamment la désambiguïsation lexicale (Véronis & 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. L'introduction d'une telle 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 de la détermination des pondérations reste délicate : elle est fortement liée 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 & Fier 2008).

De plus, les réseaux lexico-sémantiques contiennent souvent un nombre variable de types de relations, généralement ontologiques (hyperonyme) ou lexicales (synonymes, antonymes). HowNet (Dong & Dong 2006) est un exemple de telle ressource lexicale (chinois et anglais) construite manuellement et présentant un nombre important de types de relations. CYC (Lenat 1995) est un effort majeur de construction manuelle d'une base de connaissances générales basées 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 (Navigli

& Ponzetto 2010) est une extraction/compilation automatique sous forme d'un réseau multilingue de grande taille des informations de Wikipédia.

C'est à ce type de méthode, reposant sur des « jeux sérieux finalisés » ou « jeux avec finalité d'acquisition de ressources » ou encore en anglais Games With A Purpose (GWAP) (von Ahn 2006) appliqué à ce type de ressource (les réseaux lexico-sémantiques) que nous nous intéressons dans ce chapitre. Dans une première partie, nous nous interrogerons sur ce qui fait (ou pas) qu'un jeu de nature lexicale attire les joueurs et produit des ressources et quelles sont les qualités que ce jeu doit posséder. Nous aborderons également les apports du jeu du point de vue des concepteurs. Nous mentionnerons ensuite le problème de la tricherie et ses effets négatifs sur la ressource produite et la perte d'intérêt du jeu, avant d'examiner les intérêts de l'aspect communautaire d'un jeu par consensus populaire. Enfin, nous illustrerons notre propos en détaillant l'expérience du projet JeuxDeMots (noté JDM par la suite) (Lafourcade 2007) qui vise à produire un réseau lexico-sémantique. De plus, nous présenterons AKI, un des jeux du projet JDM, qui offre une possibilité d'évaluation qualitative des données recueillies. Nous analyserons les données collectées et nous en déduirons quelques difficultés rencontrées par les joueurs, mais également dans la conceptions des jeux.

## 2. Quelles propriétés pour un jeu lexical ?

Les caractéristiques nécessaires pour qu'un jeu lexical ait du succès peuvent se trouver dans les réponses à quelques questions simples : que peut-on demander aux joueurs ? comment ? sous quelle forme ? Certains éléments de la grille d'analyse qui suit sont transposables à des jeux non lexicaux. Peu d'analyses ont été proposées concernant les GWAP et la plupart se situe un contexte très général (von Ahn & Dabbish 2008) ou constituent plus une description de ce qui existe (Thaler et al. 2011).

### 2.1 ... 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

le cas de construction de ressources lexicales et terminologiques, la grande taille du vocabulaire offre des possibilités de renouvellement quasiment illimitées. La variété de parties proposées est encore renforcée si la ressource inclut des relations lexico-sémantiques de types variés, des niveaux de langues, des informations en rapport avec les phénomènes de polysémie, etc. L'ambiguïté inhérente au lexique est aussi une source d'intérêt et de surprise pouvant être exploitée avec profit.

- facile à appréhender

Beaucoup de joueurs potentiels ne lisent pas les consignes, il faut donc que le jeu soit intuitif. Dans le cadre d'un jeu lexical, de nombreuses relations sont directement compréhensibles par le locuteur et ne nécessitent généralement pas un bagage linguistique élaboré. Il est à peu près certain que les données nécessitant une formation linguistique poussée pour être appréhendées (par exemple, des arbres d'analyse morpho-syntaxiques) ne peuvent pas être demandées à des joueurs.

- 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 retrouve renforcée en quantité. Minimiser l'interaction entre une fin de partie et la suivante est une caractéristique à vérifier autant que possible.

- valorisation des joueurs

Un système de points et de classement entre les joueurs rend en général le jeu plus attirant : la compétition entre les joueurs fait que certains vont y consacrer beaucoup de temps, acquérant de l'expérience, ce qui les rend non seulement plus productifs mais augmente globalement la qualité des contributions. Il est très rare qu'un joueur soit repoussé par la présence de classements, et au pire choisit de l'ignorer.

Dans des jeux où il y a confrontation au système, les joueurs peuvent être mis en situation de se sentir « plus intelligents que la machine » (par exemple, dans AKI le jeu de devinettes (Lafourcade et al. 2009) présenté au paragraphe 3.3, un grand nombre de parties visent à piéger le système, ou à lui apprendre quelque chose), les joueurs ont de plus le sentiment (réel) d'être utiles dans l'acquisition des connaissances.

- 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 JDM, présenté au paragraphe 3.1, mais par la suite les joueurs préfèrent des consignes plus difficiles, au risque sinon 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, niveau 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. Toujours dans un jeu lexical, le niveau de difficulté peut être calqué sur les propriétés lexicales déjà découvertes (et validées) des termes du lexique. Par exemple, il est possible de faire jouer les participants sur la distinction au niveau des associations d'idées pour un verbe donné entre sa forme transitive et intransitive.

- 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 données lexicales que l'on cherche à récolter. Les approches lexicales en général, et les réseaux lexico-sémantiques en particulier, peuvent être fondés sur des micro-tâches dont la granularité est très fine et le temps de construction très faible. Dans le cas de JDM, une enquête a montré que la durée prévue d'une session par les joueurs était de l'ordre de 10 minutes, mais que la durée effective avoisinait 25 minutes.

- 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.

## 2.2 ... pour les concepteurs

Pour les concepteurs d'un jeu dont l'obtention de ressources de qualité constitue le but initial, cette méthode d'acquisition présente un certain nombre de qualités :

- les ressources produites sont le fruit de résultats consensuels, partagés par un grand nombre de joueurs ;
- le résultat est la constitution d'une base de connaissances générales communes, auxquelles viennent nécessairement s'ajouter des connaissances plus spécifiques ;
- 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 1000 occurrences de relations lexicales par jour dans le cadre du projet JDM). Le niveau d'éducation des joueurs est très majoritairement bac+2 ou plus ;
- un jeu d'acquisition lexicale doit si possible s'adresser à une partie aussi large que possible de la population. Cependant, la nature même des objets manipulés peut réduire fortement cette espérance. Dans le cas de JDM, 70 % des joueurs ont entre 30 et 60 ans, et 60 % sont des femmes, et la quasi-totalité a un niveau d'étude supérieur ou égal à bac+2 ;
- on remarquera enfin que d'un point de vue éthique la sollicitation de contributeurs à travers des jeux semble ne pas poser autant de problème que dans un cadre rémunéré (Sagot et al. 2011), même si à première vue les tâches semblent très similaires. La question de l'efficacité est également débattue, mais *a priori* la motivation des participants à bien jouer induit la constitution d'une ressource de bonne qualité. Ce dernier point ne va pas de soit et mériterait d'être évalué plus avant.

### 2.3 La question du roulement

Le roulement (renouvellement des joueurs) peut être très court, d'autant plus s'il y a phénomène de mode au sein d'un sous-groupe ou d'une petite communauté de joueurs. Selon notre expérience sur JDM, le roulement est de quelques trois semaines. Nous avons mesuré cette durée de façon continue après chaque campagne d'information sur JDM, que ce soit lors de conférences scientifiques ou auprès d'étudiants.

Nous avons constaté un autre phénomène, à savoir qu'un certain nombre de joueurs « reviennent », après un longue période d'inactivité, parfois simplement pour jouer à nouveau, mais le plus souvent avec une attitude différente, plus ciblée (par exemple, dans AKI pour vérifier si le système connaît telle ou telle information).

Un roulement des joueurs, relativement court, est principalement un inconvénient : la majorité des joueurs n'a pas le temps d'acquérir suffisamment d'expérience dans le jeu pour en ressentir toutes les subtilités et en reste trop souvent à un niveau sommaire. Indiquons toutefois que le côté addictif d'un jeu, ainsi que le système de points et de classement qui favorise l'émulation entre les joueurs, ont pour conséquence qu'une proportion non négligeable de joueurs permettent d'acquérir des ressources qui s'apparentent plus à des ressources 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.

#### 2.4 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) : elle conduit à la création de bonnes ressources, voire ponctuellement meilleures, mais induit un sentiment d'écœurement chez les autres joueurs, et donc une certaine désaffection (s'il y a des joueurs qui trichent, cela fausse la compétition, et de plus le caractère sérieux du jeu s'en trouve grandement érodé).

La réponse à donner nous semble être celle du débusquage des zones sensibles à l'aide de joueurs pilotes, de préférence compétents informatiquement. En effet, dans le cadre de JDM, de nombreuses attaques, par injections de codes informatiques notamment, ont eu lieu. En général, cette phase n'est critique qu'aux premières heures d'un tel projet.

Le second type de tricherie, totalement négative, vise à « pourrir » la base et relève d'avantage 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. Cette tricherie peut être fastidieuse à contrôler manuellement (par un expert) ; elle peut être délicate, voire quasi impossible, à détecter automatiquement. En effet, une information originale peut aussi être censée et pertinente.

La solution mise en œuvre dans JDM consiste à jouer les parties par paires de joueurs, en asynchrone et de façon anonyme. On ne mémorise que l'intersection des réponses de ces deux joueurs : si l'un d'eux a répondu des inepties, elles ne seront pas mémorisées et le fait d'afficher le résultat découragera le joueur malveillant. Si une activité suffisante est maintenue au sein du jeu, la probabilité que deux joueurs vandales se rencontrent reste faible.

#### 2.5 Intérêts de l'aspect communautaire

Une partie non négligeable de la motivation des joueurs repose sur l'aspect communautaire d'un jeu. Cet aspect réside dans l'attrait général que suscitent les jeux communautaires (le joueur fait partie d'un groupe) et plus généralement dans les réseaux sociaux. Pour la pérennité d'un jeu, son caractère communautaire présente plusieurs avantages :

- à jouer tout seul dans son coin, le joueur se lasse vite ;
- avec la création d'un forum, les joueurs peuvent discuter entre eux, échanger leurs points de vue, contester (éventuellement approuver) les opinions des autres : les joueurs deviennent de plus en plus experts (ex. : dans JDM la différenciation entre les notions de *consituant* et de *matière*<sup>1</sup>). De plus, les joueurs fournissent un retour sur le jeu et ses modalités, important pour les concepteurs ;
- il peut y avoir une forme d'appropriation du jeu par les joueurs (ex. : un joueur peut décider de jouer sur un thème qui l'intéresse (« je suis le spécialiste de ce domaine »), les autres thèmes seront renseignés par d'autres joueurs).

### 3. Les jeux et contre-jeux du projet JDM

La présentation et l'analyse de certaines propriétés des jeux du projet JeuxDeMots, permettent de rendre concrètes certaines des propriétés désirables mentionnées ci-dessus. De plus, l'expérience menée depuis plus de 5 ans permet d'évaluer l'impact effectif en termes de quantité et qualité des ressources collectées.

Dans le cadre du projet JDM, décrit en particulier dans (Lafourcade 2007, Lafourcade & Joubert 2010), nous avons développé plusieurs jeux, avec chaque fois des objectifs différents : pour chacun, certaines limites sont évoquées. Nous ferons remarquer que chaque jeu est pourvoyeur d'informations mémorisées dans le réseau lexical selon des modalités bien précises, mais constitue également un contre-jeu<sup>2</sup> permettant de compenser au moins partiellement les biais induits par chacune des autres modalités.

#### 3.1 JeuxDeMots

JDM<sup>3</sup> : le tout premier jeu du projet, mis en service en juillet 2007 a comme but la constitution du réseau lexical. La base lexicale a été amorcée avec environ 150000 termes et aucune relation entre ces termes. Au bout de cinq ans au 1er janvier 2013, le réseau contient 250 000 termes et plus de 1,5 millions de relations.

---

1. Par exemple : le *ped* constitue un constituant de la *chaise*, alors que le *bois* en est une matière possible.

2. Merci à Pierre Zweigenbaum pour avoir trouvé le terme particulièrement pertinent de « contre-jeu ».

3. JeuxDeMots est accessible à l'adresse <http://jeuxdemots.org>

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 en 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 1 présente une partie typique de JeuxDeMots où il est demandé d'énumérer ce que peut évoquer le terme jeu sérieux. Le joueur a fait 16 propositions et, pour certaines d'entre elles, a sélectionné l'usage qui lui semblait être pertinent. Par exemple, parmi les usages possibles de *jouer*, il a choisi *jouer (s'amuser)*.

Le résultat de la partie est présenté dans la Figure 2. La liste des termes en commun avec l'autre joueur est présentée. Chacun des deux joueurs récolte les gains annoncés. Le score étant élevé, un certain nombre d'actions sont disponibles et constituent une récompense supplémentaire au joueur.

Une partie ratée peut être l'objet d'un procès de la part de celui qui vient de la jouer. Un procès consiste à rendre publiques les réponses des deux joueurs et exposer les griefs du plaignant. Les autres participants ont alors l'occasion de s'exprimer et de voter pour ou contre la partie mise en accusation. À l'issue d'une période de temps (d'une semaine) le procès est clos et le gagnant (l'accusateur ou l'accusé) reçoit un dédommagement symbolique. 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, à savoir développer une réputation positive en offrant des parties intéressantes. Le développement du réseau lexical s'en trouve accéléré et bonifié, l'énorme majorité des joueurs s'abstenant d'offrir des couples terme/relation sans intérêt. L'analyse comportementale des joueurs offrant des cadeaux montre que la pression sociale (via le forum ou les procès) empêche les comportements stériles ou d'anti-jeu.

JeuxDeMots est clairement un jeu de coopération/compétition. À bien jouer, on augmente son propre score global mais aussi ceux des autres joueurs. La compétition est induite par la présence de classement sur plusieurs critères (points gagnés, nombre de parties jouées, efficacité etc.) et renforce l'activité et l'application de chacun des joueurs à jouer le mieux possible.

Les autres (contre-)jeux sont dits « fermés » dans la mesure où une ou plusieurs propositions sont affichées et doivent être attribuées ou validées. De façon générale, les parties sont plus courtes que dans JDM et sont jouables via des



Figure 1. Exemple d'une partie de JeuxDeMots où le joueur doit indiquer à quoi lui fait penser le terme « jeu sérieux ».



Figure 2. Résultat de partie de JeuxDeMots.

Smartphones ou des tablettes tactiles. Ce dernier point n'est pas qu'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 AKI reste toutefois un jeu partiellement ouvert, car si le système ne trouve pas la bonne réponse, le joueur est invité à faire une proposition.

### 3.2 PtiClic

Le jeu PtiClic<sup>4</sup> (Lafourcade & Zampa 2009), lancé un an après JDM, a pour but la consolidation du réseau créé grâce à JDM : il permet de renforcer ou propager les relations existantes. PtiClic est un jeu fermé d'attributions par glisser-déposer : le joueur doit déplacer des termes dans les catégories qui lui semblent pertinentes relativement à un mot cible. Par exemple (Figure 3), le mot cible étant *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).

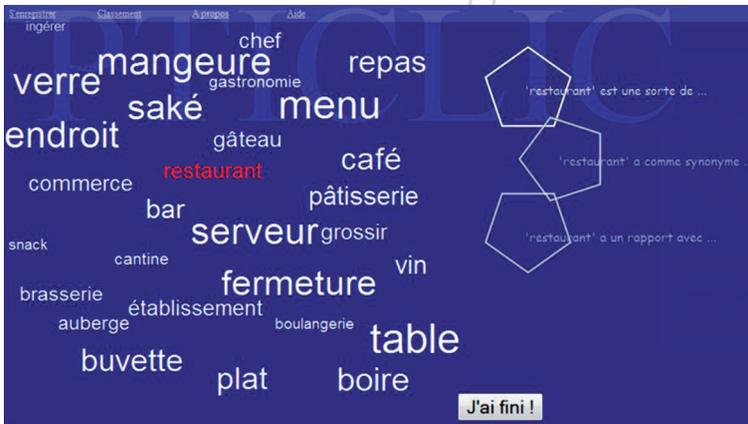


Figure 3. Premier exemple de deux parties typiques de PtiClic. Le mot cible est *restaurant* et les trois relations cibles l'hyperonymie, la synonymie et l'association libre.

Une partie de PtiClic est d'une durée non limitée mais en pratique inférieure à celle de JDM (entre 30 et 45 secondes), mais supérieure aux jeux à clics. La modalité d'interaction rend possible l'utilisation de tablettes ou de smartphones. Au premier janvier 2013, environ 60000 parties de PtiClic ont été jouées depuis son lancement.

L'effet de PtiClic est de faire migrer des relations de la relation la plus générale d'idées associées vers des relations plus spécifiques. Les relations visées sont l'hyperonymie, l'hyponymie, partie/tout etc. Il s'agit donc de relations ontologiques. Les relations strictement lexicales (termes de la même famille) sont dans le cadre de ce jeu plus difficiles à sélectionner avec une certitude suffisante de façon à créer des parties intéressantes. L'activité de PtiClic n'est pas de nature correctrice.

4. PtiClic est accessible à l'adresse <http://www.lirmm.fr/pticlic/pticlic.php>



Figure 4. Deuxième exemple de deux parties typiques de PtiClic.

La relation idées associées reste proposée comme cible possible lors des parties. En effet, certains termes du nuage sont extraits via des méthodes basées sur LSA ou dans un voisinage supérieur à 1 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 & 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.

### 3.3 Jeux à devinettes

AKI<sup>5</sup> est un jeu lancé début 2010 qui peut être considéré comme un jeu de devinettes, mais également comme outil d'aide au problème du Mot sur le Bout de la Langue (MBL). L'objectif de la conception d'AKI était de d'obtenir une modalité permettant une évaluation qualitative du réseau lexical JDM. Cependant, le succès de ce jeu et le fait que beaucoup de joueurs cherchent à « piéger » AKI, ont également conduit à un enrichissement du 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, *:isa animal*).

5. AKI est accessible à l'adresse <http://www.lirmm.fr/jeuxdemots/AKI.php>



Figure 5. Deux exemples de partie de AKI avec une réussite (*médoc*) et un échec (*gare de triage*).

L'évaluation du réseau via AKI a été présentée dans (Lafourcade et al. 2011). Le résultat le plus notable est que AKI 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, AKI offre aussi un renforcement du réseau et des détections d'incohérence par un mécanisme d'apprentissage. Depuis son lancement, plus de 23000 parties de AKI ont été jouées (l'enregistrement à des fins statistiques n'ayant été réalisé qu'à partir de la 10000e).

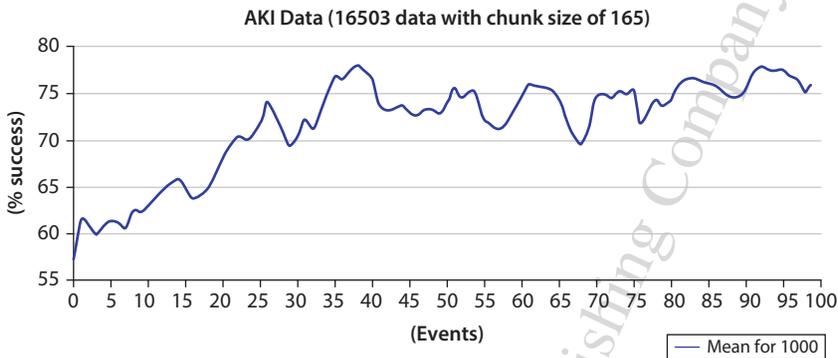


Figure 6. Évolution du taux de réussite d'AKI à deviner correctement le terme ciblé par le joueur.

### 3.4 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 proposées (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 terme de modification de la ressource lexicale. La présentation suite au choix effectué par l'utilisateur de sa position dans la répartition des votes (par exemple minoritaire ou majoritaire) est la réponse du jeu et forme la récompense qu'attend le joueur. L'objectif implicite du jeu, implicite car non annoncé explicitement aux joueurs, est de répondre honnêtement 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 car il ne sont pas sujets au forçage de réponses et que les joueurs sont relativement peu indulgents avec les réponses limites.

**AskIt<sup>6</sup>** : est un jeu à clics 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 « je ne sais pas », il permet un enrichissement et une consolidation du réseau, en particulier en ce qui concerne les raffinements des termes polysémiques.

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



Figures 7 et 8. Exemples de deux écrans consécutifs de AskIt.

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'appartenir ou non à la majorité des réponses. Le gain symbolique en points s'accumulant en une barre d'état constituant un objectif (la remplir) est une incitation psychologique forte à jouer honnêtement. Le taux de trollisme (joueurs répondant volontairement quelque chose d'erroné) a été estimé à moins de 1 % de l'ensemble de plus de 300000 parties jouées à ce jour.

L'algorithme de sélection des questions est double :

- sélection par triangulation dans le réseau ;

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

$A r1 B r2 C \Rightarrow ? A r3 C$

avec des sélections particulières de triplet  $(r1, r2, r3)$ . Par exemple, si *canari* isa *oiseau* et *oiseau* agent *chanter* alors avons nous *canari* agent *chanter* ? Nous ferons remarquer que les réponses à ce genre de question sont loin d'être systématiquement positives. Par exemple, un contre exemple serait *poule* isa *oiseau* agent *chanter*  $\Rightarrow^*$  *poule* agent *chanter*. 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* (*oiseau*) raffinement *frégate* partof *coque*  
 $\Rightarrow^*$  *frégate* (*oiseau*) partof *coque*

*frégate* (*navire*) raffinement *frégate* partof *coque*  
 $\Rightarrow$  *frégate* (*navire*) partof *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 que 40 % des propositions faites aux joueurs sont erronées. Nous ferons remarquer que cela est désirable et volontaire. En effet, un algorithme « trop bon » donnant une majorité de réponses positives, serait ennuyeux pour le joueur. La proportion importante de propositions fausses 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 » baisse légèrement le poids de la relation en jeu. Certaines relations peuvent donc devenir négatives (dites inhibitrices). Par exemple, *autruche* agent *voler* a un point négatif (de -45 dans la base au moment de l'écriture de ce chapitre). Ces relations sont particulièrement intéressantes en analyse sémantique par propagation de contraintes sur un réseau d'occurrences de termes, en ce qu'elles jouent le rôle d'interdictions. Environ 10000 relations inhibitrices ont été obtenues avec AskIt depuis 12 mois.

AskIt possède un certain nombre de caractéristiques notables :

- c'est un jeu addictif : le système « relance » le joueur qui arrêtera le jeu sur une question simple qu'il a eu le temps de lire (env. 2 à 3 secondes) et dont il connaît en général la réponse. Sentiment de satisfaction (on a fini une partie) / frustration (on arrête sur une partie en cours à laquelle on peut répondre d'un simple clic) ;
- il peut être amusant, voire accrocheur, en particulier sur les termes polysémiques (ex. : *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 ne sont pas linguistes ? On obtient une connaissance générale commune et non une connaissance de spécialistes. Attention : si la question est trop complexe ou ambiguë, cela peut entraîner un grand nombre de non-réponses ;
- le résultat (qu'en ont pensé les autres joueurs ?) est donné immédiatement : il y a valorisation des joueurs par comparaison avec la réponse majoritaire (« je suis comme les autres », quand on a hésité / « je suis plus malin qu'eux », si on a fourni une réponse astucieuse ou réfléchie).

**LikeIt**<sup>7</sup> : ce jeu 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 permet un enrichissement du réseau en termes de polarité. 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. Comparativement aux qualités énumérées dans les sections précédentes, LikeIt possède un certain nombre de caractéristiques notables :

- comme Askit, LikeIt est un jeu addictif, à parties courtes et relance immédiate ;
- les réponses sont très simples : positive / négative / neutre, le joueur n'ayant pas l'impression de renseigner la base. La variété du vocabulaire fait que le joueur n'a plutôt pas tendance à voir ce jeu comme un sondage ;
- certains termes sont très partagés (ex. : 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 dans lesquelles il joue (ex. : le terme *baccalauréat* présente un sentiment négatif quand on est lycéen, mais beaucoup plus positif lorsqu'on l'a obtenu).

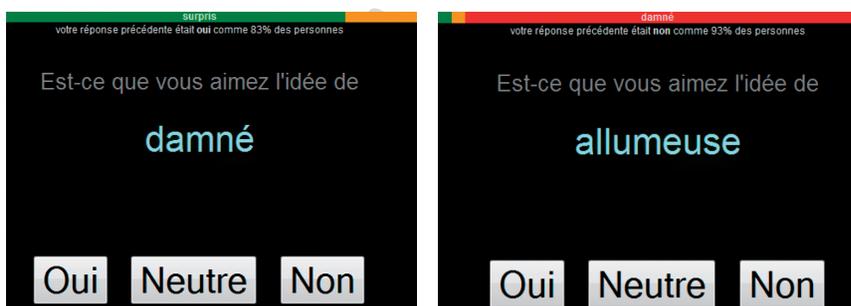


Figure 9. Exemple de deux écrans consécutifs de LikeIt. La réponse donnée dans l'écran de gauche se trouve en haut de l'écran de droite, fournissant ainsi un retour direct au joueur avec une relance immédiate.

La sélection d'un terme à proposer au joueur est effectuée pseudo-aléatoirement dans le réseau lexical de JeuxDeMots. Un terme déjà polarisé partiellement positivement ou négativement (mais pas uniquement neutre) est sélectionné au hasard

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

et un des termes qui lui est relié est sélectionné. Ainsi, cet algorithme simplissime (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 diminuant ainsi l'intérêt du jeu. En trois mois, plus de 24000 termes ont été polarisés pour un total dépassant 150000 clics.

#### 4. Analyse des données collectées et propriétés des jeux

Il est possible d'analyser les données du réseau JeuxDeMots et d'en déduire certaines difficultés qu'éprouvent les joueurs face à certains types de relations lexico-sémantiques.

##### 4.1 Quelques chiffres

Depuis le lancement du projet JeuxDeMots, le réseau lexical a acquis environ 1.5 million de relations. La base initiale de 152000 termes dépasse actuellement 250000 (au 1er janvier 2013). Il y a environ 60 relations lexico-sémantiques dans JDM, la plupart d'entre-elles pouvant être jouées.

**Tableau 1.** Nombre d'occurrences des principales relations du réseau lexical obtenu avec JeuxDeMots.

696377 associations libres	86118 r_domain	165594 synonymes
52596 hyperonymes	12877 antonymes	13423 hyponymes
15709 parties (meronymes)	11015 tout (holonymes)	13814 locution
13202 agent (verbe)	10365 patient (verbe)	20194 lieux typiques
8507 instrument (verbe)	15685 caractéristiques typiques	

Certaines relations ne sont pas jouables, mais font partie du réseau lexical :

- 88758 catégories morphosyntaxiques car jouer sur les parties du discours serait sans intérêt ;
- 65041 lemmes qui ont été obtenus de façon automatique ou contributive (hors jeux).

En effet, ces informations ne sont pas amusantes à jouer et présentent trop peu de variations dans une entrée pour présenter un intérêt ludique. Il est donc nécessaire de les entrer manuellement ou par tout autre moyen (semi) automatique.

- 11708 raffinements sémantiques issus du calcul de cliques (Lafourcade & Joubert 2010)

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 des termes représentatifs des usages possibles. Les joueurs trouvent le glosage particulièrement difficile mais intéressant. Parfois, l'introduction de périphrases semble être la seule réponse possible si aucun hyperonyme ou synonyme ne semble convenir.

#### 4.2 Quelques relations difficiles

Enfin, certaines relations posent plus de problèmes aux joueurs :

- cause / conséquence et producteur / produit

La distinction du choix du bon membre de chacun des couples de relations semble poser problème aux nouveaux joueurs. Cependant, l'analyse de certains fils de discussion liés au projet JDM, montre que le problème n'est pas d'ordre linguistique, mais plus lié à la connaissance du monde. Par exemple, entre *fièvre* et *malade*, l'identification de la cause ou de la conséquence semble difficile.

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

Les hyponymes semblent chez beaucoup de joueurs inclure les instances (mais pas l'inverse). Par exemple, *Jolly Jumper* est vu aussi bien comme une instance de *cheval* qu'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 est le ou les verbes de construction de l'entité/artéfact (par exemple, *construire* pour *maison*, ou *écrire/éditer* pour *livre*) (Pustejovsky 1993). L'implication agentive est le ou les verbes présidant à la construction (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 s'ensuit un nombre important de parties erronées.

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

L'axe d'intensité pour un terme est souvent parfaitement subjectif aux joueurs, et les plus intenses (dont un exemple canonique est *fièvre magn forte fièvre*, *fièvre de cheval*, etc.) incluent les mélioratifs (*repas mélioratif excellent repas*).

- les sentiments/émotions

La notion de sentiment/émotion semble parfois très large, et il arrive que des termes a priori étrangers aux sentiments soient utilisés comme tels. Par exemple, *or* (*argent*) a pour sentiment *richesse*, ce qui semble curieux mais toutefois pouvant être interprété comme associé au *sentiment de richesse*.

- termes de la même famille lexicale

Il s'agit avec cette relation de lister des termes appartenant à la même famille lexicale de façon dérivationnelle. Il arrive que les résultats soient étonnants, par exemple pour le substantif *nuit* nous obtenons : *nuitée, nocturne, noctambule, nuitamment, minuit, noctambulisme, chemise de nuit, noctuelle, nuitement, nuité, noctule, bonnet de nuit, nuire, nuisible*. Manifestement *nuire* et *nuisible* sont des intrus. Une évaluation par échantillonnage semble indiquer qu'environ 5 % des termes associés par cette relation constituent des erreurs (erreurs parfois largement partagées parmi les locuteurs).

- problème avec la typicalité

Certaines relations (comme agent, patient, caractéristique, etc.) demandent a priori des associations typiques. Par exemple *rugir* est typique d'un *lion*, mais pas *courir* bien que possible. De même, *tigre* est typiquement *féroce*, mais peut aussi être *vieux*. Nous constatons à la vue des données que l'ensemble des possibles a tendance à apparaître dans le réseau, mais avec des pondérations beaucoup plus fortes pour les associations typiques. C'est une bonne nouvelle pour ces relations qui deviennent plus couvrantes que ce qui avait été espéré.

#### 4.3 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, on peut légitimement se poser les questions : qu'est-ce qu'une réponse erronée ? et comment la détecter ? De plus, cette pénalisation serait probablement vécue négativement par le joueur, pourrait le dissuader en cas d'hésitation, voire induire une démotivation et une désaffection pour le jeu.

Un des biais majeurs de JDM veut que la consigne soit globalement forcée – « il faut absolument trouver et mettre des réponses ». Le joueur a la possibilité d'indiquer qu'il n'en trouve pas, mais il arrive qu'il préfère quand même indiquer ce qui lui semble le plus approchant. C'est ainsi que pour des spécifiques de *semaine* ont été proposés les jours de la semaine : *lundi, mardi*, etc. Les relations erronées relatives à ce type de phénomène de forçage ont été évaluées (par échantillonnage) à environ 0.1 % des relations. Il est à noter que la présence des contre-jeux vient contrebalancer largement ce biais propre aux jeux ouverts.

Ceci étant dit, nous avons remarqué que les espaces d'échanges entre joueurs (les procès et le forum) sont des lieux privilégiés pour éclaircir la sémantique des relations quand elle n'est pas claire au joueur. Ainsi, à mesure de sa pratique, ce dernier a tendance à augmenter son expertise.

De façon plus précise nous avons évalué les performances globales des joueurs. Dans JDM, chaque partie en stock est signée (nous savons quel joueur en est le créateur) et de plus on peut calculer de façon incrémentale le nombre moyen de termes en commun entre les joueurs deux à deux (on parlera d'efficacité). De plus, avec le concours de joueurs volontaires, nous avons une idée partielle de leur profil (âge, sexe, niveau d'étude, domaines de compétence, etc.).

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. 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. A l'analyse de leurs parties, les « linguistes » font moins de propositions que la majorité des autres joueurs, et leurs propositions sont majoritairement des termes les plus activés du réseau. Donc, la plupart du temps, il s'agit de propositions correctes mais peu imaginatives et n'alimentant pas la longue traine.

A l'inverse, les joueurs gravitant autour du FLE (comme enseignants) ont tendance à être meilleurs que les joueurs normaux et a fortiori les linguistes. Leur efficacité moyenne est d'environ 3,4. De plus, il y a au moins 3 FLE dans les premiers joueurs de JDM concernant le score d'efficacité.

Les joueurs ayant une activité professionnelle relevant de la langue (linguistes, FLE, TAL, etc.) représentent environ 10 % des joueurs, et 7 % 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 discutables. Environ 0,4 % des relations erronées sont repérables automatiquement. Il s'agit par exemple des double hyperonymes (A hyper B et B hyper 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 AKI a montré que sur un jeu de devinettes le système (strictement fondé sur le réseau de JeuxDeMots) trouvait la bonne réponse dans 75 % des cas (à comparer à 48 % pour des devineurs humains). Les données du réseau lexical JDM sont en accès libre et sont libres de droits. Elles peuvent être récupérées globalement sur la base d'une exportation mensuelle<sup>8</sup> ou dynamiquement terme par terme<sup>9</sup>.

8. <http://www.lirmm.fr/~lafourcade/JDM-LEXICALNET-FR/?C=M;O=D>

9. <http://www.jeuxdemots.org/rezo-xml.php>  
© 2013. John Benjamins Publishing Company  
All rights reserved

#### 4.4 Deux exemples d'analyse problématiques

Le jeu **SexIt** 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, que les comportements ou pratiques sexuelles, ou les 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'un certain nombre de parties (environ une vingtaine) un profilage humoristique de type complètement obsédé/oie blanche. Ce profilage n'avait aucune prétention psychologique mais simplement d'augmenter le côté ludique du jeu. Le calcul du profil était relativement trivial : on était qualifié d'obsédé (resp. de oie blanche) si on avait une majorité de réponses de type *sex* (resp. *no sex*) pour des termes que les autres joueurs trouvaient majoritairement *no sex* (resp. *sex*).



Figure 10. Exemple de deux écrans consécutifs de **SexIt**. La réponse donnée dans l'écran de gauche se trouve en haut de l'écran de droite, fournissant ainsi un retour direct au joueur avec une relance immédiate.

Il l'issue d'environ 5000 parties sur environ 500 termes, 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 pour *sex*. De façon similaire, des termes clairement sexuels avaient environ un tiers de votes *no sex*. Ce pourcentage de votes inattendus était bien trop élevé pour être attribuable à une variation naturelle due aux erreurs ou aux quelques joueurs pratiquant le trollisme.

L'analyse comportementale des joueurs test a montré qu'environ un tiers d'entre eux avait tendance à choisir leurs réponses en fonction de ce qu'il souhaitait obtenir comme qualificatif, à savoir obsédé ou oie blanche. Le choix n'était plus du tout fondé sur l'appréciation personnelle mais sur un objectif à atteindre a priori. Atteindre cet objectif est immédiat en cliquant systématiquement un seul des deux choix.

La mise à jour de SexIt par la suppression des mentions (avec la réinitialisation des données) a permis de régler immédiatement le problème. Dans les données obtenues dans 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 était contravariante avec la qualité des données produites. La réalisation des objectifs possibles par le joueur doit impliquer la création de ressources de qualité c'est-à-dire reflétant fidèlement sa pratique langagière. Les autres objectifs pouvant potentiellement biaiser les données doivent être neutralisés en rendant le jeu inintéressant ou dévalorisant.

Un autre exemple est donné par le jeu Verbosity (von Ahn et al. 2006b) qui vise à recueillir des connaissances d'ordre général via un jeu de devinettes. Ce jeu ne fait pas partie du projet JDM, mais il permet également l'acquisition de ressources de nature lexico-sémantique. A l'inverse d'AKI où un joueur se confronte à la machine, ici deux joueurs sont impliqués. Les joueurs sont tour à tour fournisseur d'indices et devineur. Le mot à faire deviner est sélectionné par le système (nous rappelons que dans AKI le joueur choisit le terme à faire deviner à la machine). Le jeu fonctionne en anglais.

L'interface utilisateur propose des zones de saisie de texte typé sémantiquement que le joueur est invité à renseigner de façon à faire deviner le mot choisi par le système. L'autre joueur voit donc s'afficher les indices typés et peut faire des suggestions. Le joueur ne peut entrer comme indices que des termes connus du système.

Un des problèmes majeurs de l'interaction proposée est que le joueur peut mettre autant de mots qu'il veut, et le plus souvent ne tient pas du tout compte du typage. Le jeu se résume donc en pratique à donner une série de termes librement associés au terme cible. De plus, souvent une forme de discussion spontanée s'amorce entre les deux joueurs, dont un exemple prototypique (en français) serait :

joueur poseur (P)	<i>couleur</i>
joueur devineur (D)	<i>rouge</i>
P	<i>inverse</i>
D	<i>vert</i>
trouvé !	

Un autre type de partie *corrompue* est :

www.gwap.com/verbosity-a/verbosity.html?9656174&228584&guest228584&0

score 0 time 3:35

Verbosity  
it's common sense.

it is used for poking. 325 pts!

your partner's clues  
it looks like horn.  
it is a type of pointy thing.  
it has rose protection.

guess the secret word

new guess...  
+ submit  
→ pass

BONUS!  
5,000 PTS

Most Points Today

1	Jeff	275 k
2	Pleasidobby	264 k
3	DaftyMcDaff	30 k
4	yul511	25 k
5	Lotte	15 k
6	guest228573	12 k
7	guest228550	11 k
8	guest228552	9,250
9	is	2,750
10	guest228563	2,750

Figure 11. Exemple de partie de Verbose pour le joueur qui doit deviner le mot à partir des indices proposés.

the secret word is... squirrel. 225 pts!

clues

it is cute

it is a type of animal

it has tail fur

it looks like

it is typically near nuts

it is the opposite of + submit

→ pass

guesses

BONUS!  
5,000 PTS

Figure 12. Exemple de partie de Verbose pour le joueur qui doit faire deviner le mot à partir des indices proposés. Les indices proposés ici suivent les consignes de types.

P            *blanc poudre froid*  
 D            *neige*  
 trouvé !

dont le premier indice sera difficilement distinguable d'un mot composé, comme par exemple :

P            *pomme de terre*  
 D            *patate*  
 P            *huile friture doré*  
 D            *frite*  
 trouvé !

Et ce, sans jamais tenir compte du type associé à la zone de texte remplie. Il est clair que dans ces conditions, les données lexicales doivent être extrêmement bruitées. Toutefois, n'ayant pu avoir accès aux données collectées, il est difficile d'avoir une estimation précise de leur qualité. Il ne semble guère possible de redresser (en réattribuant les bons types et en segmentant les indices si possible) les données ainsi collectées sans disposer d'une ressource externe.

## Conclusion

Dans ce chapitre, nous avons montré l'intérêt des jeux sérieux et plus particulièrement des jeux à finalité qui permettent d'acquérir facilement et à moindre coût des ressources lexicales de qualité. Toutefois, ceci est réalisable à condition que ces jeux possèdent un certain nombre de caractéristiques afin d'attirer et de fidéliser des joueurs qui contribuent à la création d'une ressource, généralement sans avoir conscience de leur contribution. A la conception traditionnelle d'un jeu sur internet se rajoute la difficulté d'avoir une boucle ludique covariante avec la boucle d'acquisition de données. Nous avons donné des exemples de jeux et de données issus du projet JeuxDeMots, ainsi que certains types de conception et d'interaction posant problème.

Cette méthode, relativement récente puisqu'elle n'est réalisable que depuis l'émergence d'internet, ne vient pas en opposition par rapport aux méthodes traditionnelles à savoir essentiellement manuelles qui, elles, existent depuis plusieurs siècles. 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 de n'être qu'un complément par rapport à ces outils de référence et de démontrer la faisabilité réelle d'approches fondées sur de la peuplologie ludique.

Nous avons observé que la collecte d'informations relevant aussi bien du lexique que du sens commun se faisait aisément via des jeux et avec une qualité très satisfaisante. Plus l'information demande une expertise linguistique poussée, plus les résultats sont aléatoires. Concernant la constitution d'un réseau lexical, il y a peu de données exigeant une forte compétence linguistique dépassant l'intuition du locuteur. Il n'en serait certainement pas de même, s'il s'agissait par exemple de constituer une banque d'arbres d'analyses syntaxiques, ou des alignements de traduction entre autres. Nous avons présenté plus en détail l'exemple d'un outil permettant la collecte de polarité qui est d'un abord très simple. L'activité est à granularité très fine et l'intuition des locuteurs correspond aux attendus. Il s'agit là d'un exemple typique où un travail de collecte manuelle semble inefficace aussi bien quantitativement que qualitativement.

L'évaluation de la ressource peut se faire avantageusement via une activité ludique. Celle-ci 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.

## Œuvres citées

- von Ahn, L. & Dabbish, L. (2004). Labelling Images with a Computer Game. *ACM Conference on Human Factors in Computing Systems (CHI)*. pp. 319–326, New York, NY, USA. ACM Press.
- von Ahn, L. (2006). Games with a purpose. *Computer*, 39(6): 92–94.
- von Ahn, L. & Dabbish, L. (2008). Designing Games With a Purpose. *Communication of the ACM*, 51(8): 58–67.
- von Ahn, L., Liu, R. & Blum, M. (2006). Peekaboom: a game for locating objects in images. In *Proceedings of CHI '06*, pages 55–64.
- von Ahn, L., Kedia, M. & Blum, M. (2006). Verbosity: a game for collecting common-sense facts. In *Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems*, volume 1 of Games, pages 75–78.
- Chamberlain, J., Poesio, M. & Kruschwitz, U. (2008). Phrase Detectives: A Web-based Collaborative Annotation Game. In *Proceedings of the International Conference on Semantic Systems (I-Semantics'08)*, Graz.
- Chamberlain, J., Poesio, M. & Kruschwitz, U. (2009). A demonstration of human computation using the Phrase Detectives annotation game. In *HCOMP '09: Proceedings of the ACM SIGKDD Workshop on Human Computation*, pp. 23–24.
- Chklovski, T. & Gil, Y. (2005). Improving the design of intelligent acquisition interfaces for collecting world knowledge from web contributors. In *Proceedings of K-CAP '05*, pages 35–42.
- Dong, Z. & Dong, Q. (2006). HowNet and the Computation of Meaning. *WorldScientific*, London.
- Fort, K., Adda, G. & Cohen, K. B. (2011). Amazon mechanical turk: Gold mine or coal mine? *Computational Linguistics*, 37(2).

- Hladka, B., Mirovsky, J., & Schlesinger, P. (2009). Play the language: play coreference. In *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pages 209–212.
- Fabre, C. & Bourigault, D. (2006). Extraction de relations sémantiques entre noms et verbes au-delà des liens morphologiques. In *Proc of TALN'06*, Leuven, Belgique, pages 121–126.
- Fellbaum, C. & Miller, G. (1998). *WordNet*. The MIT Press.
- Fisher L. (2009). *The Perfect Swarm: the Science of Complexity in Everyday Life*. Basic Books, 260 p.
- Joubert, A. & Lafourcade, M. (2012). A new dynamic approach for lexical networks evaluation. In *Proc of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey.
- Kuo, Y.L., Lee, J. C., Chiang, K. Y., Wang, R., Shen, E., Chan, C.W. & Yung-jen Hsu, J. (2009). Community-based game design: experiments on social games for commonsense data collection. In *Proceedings of the ACM SIGKDD Workshop on Human Computation, HCOMP'09*, pages 15–22, New York, USA.
- Lenat, D. (1995). CYC: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11): 33–38.
- Lafourcade, M. & Joubert, A. (2010). Computing Trees of Named Word Usages from a Crowdsourced Lexical Network. *Investigationes Linguisticae*, vol. XXI, pp. 39–56.
- Lafourcade, M. (2007). Making People Play for Lexical Acquisition. In *Proc. SNLP 2007, 7th Symposium on Natural Language Processing*. Pattaya, Thaïlande, 8 p.
- Lafourcade, M. & Zampa, V. (2009). PtiClic : a game for vocabulary assessment combining JeuxDeMots and LSA. In *Proc of CICLing (Conference on Intelligent text processing and Computational Linguistics)*. Mexico.
- Lafourcade, M., Joubert, A., Schwab, D. & Zock, M. (2011). Évaluation et consolidation d'un réseau lexical grâce à un assistant ludique pour le « mot sur le bout de la langue. In *Proc of TALN'11*, Montpellier, France, pp. 295–306.
- Lieberman, H., Smith, D. & Teeters, A. (2007). 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.
- Mihalcea, R. & Chklovski, T. (2003). Open MindWord Expert: Creating large annotated data collections with web users help. In *Proceedings of the EACL 2003, Workshop on Linguistically Annotated Corpora (LINC 2003)*.
- Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D. & Miller, K. J. (1990). Introduction to WordNet: an on-line lexical database. *International Journal of Lexicography*, 3 (4), pp. 235–244.
- Navigli, R. & Ponzetto, S. (2010). BabelNet: Building a very large multilingual semantic network. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala, Sweden, pp. 216–225.
- Ploux, S. & Victorri, B. (1998). Construction d'espaces sémantiques à l'aide de dictionnaires de synonymes. In *Proc. Traitement Automatique des Langues*, vol. 39/1, pp. 161–182.
- Pustejovsky, J. (1993). The generative lexicon. *Computational Linguistics*, pp. 409–441.
- Sagot, B. & Fiser, D. (2008). Construction d'un wordnet libre du français à partir de ressources multilingues. In *Proc of TALN 2008*, Avignon, France.

- Sagot, B., Fort, K., Adda, G., Mariani, J. & Lang, B. (2011). Un turc mécanique pour les ressources linguistiques : critique de la myriadisation du travail parcellisé. *Traitement Automatique des Langues Naturelles (TALN)*, Montpellier, France.
- Siorpaes, K. & Hepp, M. (2008). Games with a Purpose for the Semantic Web. *IEEE Intelligent Systems*, Vol. 23, No. 3, pp. 50–60.
- Thaler, S., Siorpaes, K., Simperl, E. & Hofer, Ch. (2011). A Survey on Games for Knowledge Acquisition. *STI Technical Report*.
- Véronis, J. & Ide, N. (1990). Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries. In *Proceedings of 13th International Conference on Computational Linguistics (COLING'90)*, vol. 2, pp. 389–394. Helsinki.
- Vossen, P. (1998). EuroWordNet: A Multilingual Database with Lexical Semantic Networks. Kluwer, Dordrecht, The Netherlands.
- Zampa, V. & Lafourcade, M. (2010). PtiClic et PtiClic-kids : jeux avec les mots permettant une double acquisition. In *Proc TICE2010, 7e colloque TICE*, Nancy.
- Zesch, T. & Gurevych, I. (2009). Wisdom of crowds versus wisdom of linguists measuring the semantic relatedness of words. *Natural Language Engineering*, Cambridge University Press, pp. 25–59.

## Résumé

L'acquisition de ressources lexicales est un problème récurrent en TALN. Celle-ci peut se faire automatiquement, généralement à partir de textes, mais dans ce cas elle peut être biaisée par le corpus choisi, aussi bien que par la méthode d'acquisition qui peut favoriser l'émergence de certaines informations au détriment d'autres. L'acquisition de ressources peut également être effectuée selon la tradition lexicographique, à savoir de façon manuelle. Plusieurs méthodes sont alors possibles, les ressources pouvant être créées : (a) par des experts : c'est un processus long et coûteux, souvent lourd à faire évoluer ; (b) par des approches contributives sollicitant des participants non nécessairement linguistes de formation ; (c) enfin par une stratégie plus récente consistant à mettre en œuvre des jeux accessibles sur Internet. Ces jeux permettent d'acquérir des ressources par consensus populaire entre locuteurs natifs. C'est un processus rapide, peu coûteux, très dynamique, mais cependant la question de la qualité de la ressource peut légitimement être posée.

Toutefois, il s'avère, en général, qu'une estimation donnée par un groupe d'individus non experts qui n'interagissent pas entre eux, est souvent meilleure que les choix proposés par un groupe d'experts élaborant collectivement leurs réponses. Par ailleurs, un avis moyen obtenu par un grand nombre d'individus a plus de chance de se rapprocher de l'optimum (ou d'être représentatif) que par un groupe de plus petite taille. Les jeux en ligne visant à construire des ressources exploitent implicitement ces considérations, les joueurs menant une activité isolée. Cette activité répétée un grand nombre de fois avec un grand nombre de personnes permet de faire émerger des données de qualité. Par ailleurs, l'approche ludique est une condition sans doute essentielle à l'obtention d'un grand nombre de participants motivés.

## Summary – Advantages and limits of lexical acquisition through the JeuxDeMots experiment

Lexical resources acquisition is a recurrent problem in NLP. This can be processed automatically, generally from texts, but in this case it may be biased by the chosen corpus, as well as by the acquisition methods which may favour emergence of certain information to the detriment of others. Acquisition of resources can also be made according to the lexicographical tradition, in a manual way. Several methods are thus possible, the resources can be created: (a) by experts: it is a long, expensive and often heavy process to be developed; (b) by contributory approaches which request participants not inevitably linguists; (c) last, by a more recent strategy which consists in implementing on-line games. These games allow to acquire resources by popular consensus between native speakers. It is a fast, little expensive and very dynamic process, but however the question of the quality of the resource may be asked.

However, it generally turns out that an estimation given by a group of non expert individuals who do not interact between them, is often better than choices proposed by a group of experts collectively elaborating their answers. Besides, an average opinion obtained by a large number of individuals has more chance to get closer to the optimum (or to be representative) than given by a smaller group. On-line games aiming at building resources make use of implicitly these considerations, players doing an isolated activity. This activity being repeated a large number of times, with a large number of persons, it allows to build high quality data. Besides, the playful approach is an essential condition to obtain a large number of motivated participants.