

Amener des utilisateurs à créer et évaluer des paraphrases par le jeu

Houda Bouamor, Aurélien Max, Anne Vilnat

{prenom.nom}@limsi.fr

LIMSI-CNRS et Université Paris-Sud 11 Orsay, France

Résumé. Dans cet article, nous présentons une application sur le web pour l’acquisition de paraphrases phrastiques et sous-phrastiques sous forme de jeu. L’application permet l’acquisition à la fois de paraphrases et de jugements humains multiples sur ces paraphrases, ce qui constitue des données particulièrement utiles pour les applications du TAL basées sur les phénomènes paraphrastiques.

Abstract. In this article, we present a web application presented as a game for acquiring sentential and phrasal paraphrases. It can be used both to acquire paraphrases and important quantities of human evaluations of their quality. These are particularly useful for NLP applications relying on paraphrasing.

Mots-clés : Paraphrase, acquisition de données, évaluation de données.

Keywords: Paraphrasing, data acquisition, data evaluation.

1 Introduction

L’acquisition de données est un des problèmes majeurs du TAL. La création manuelle de ressources requiert un effort considérable en termes de financement et de suivi de l’annotation. L’acquisition automatique de données, telles que des bases terminologiques, requiert des algorithmes difficiles à concevoir, parfois des données annotées permettant d’amorcer un apprentissage, et une validation des données obtenues. Le faible volume de données annotées de façon fiable est souvent un frein à la recherche expérimentale. Une solution essayée dans le domaine du TAL consiste à formuler le problème d’acquisition et d’évaluation de données sous forme d’un jeu auquel des utilisateurs seront prêts à prendre part. Ainsi, par exemple, le site web JeuxDeMots (Lafourcade, 2008), qui réalise une acquisition de relations entre termes, connaît un certain succès. Le domaine de la reconnaissance et de la génération de paraphrases manque cruellement de données, et les moyens d’évaluation actuellement utilisés se basent soit sur des procédures d’évaluations humaines spécifiques et coûteuses en temps (Max, 2008), soit sur des corpus de référence extrêmement restreints (Cohn *et al.*, 2008).

Dans cet article, nous présentons une application web basée sur un jeu visant à obtenir d’utilisateurs non seulement des paraphrases au niveau phrastique et sous-phrastique, mais également leur évaluation, et ce à grand volume. Le mode principal du jeu se présente sous forme d’une amélioration collaborative et compétitive d’articles ou de parties d’articles de l’encyclopédie en ligne Wikipedia. Dans un premier temps, chaque phrase doit être réécrite par chaque joueur inscrit à une partie. Dans un deuxième temps, les joueurs évaluent les paraphrases et les valident

de plusieurs manières possibles. Un jugement consensuel positif sur des parties de phrases proposées par un joueur mène à l’attribution de points pour celui-ci. Le résultat d’une partie est double : du point de vue utilisateur, les joueurs ont proposé des reformulations qui leur semblaient appropriées, et ont récolté un nombre de points dépendant du jugement collectif de leurs propositions ; du point de vue administrateur, des paraphrases ont été collectées et associées à un jugement humain sur leur qualité dans un contexte particulier.¹

2 Acquisition de paraphrases phrastiques

La première phase permet au joueur de s’inscrire à une partie correspondant à un article ou à une partie d’article spécifique puis de proposer, pendant une période fixée, des reformulations pour chacune des phrases constituant le texte. L’interface permet une édition phrase à phrase dans des zones textuelles indépendantes. L’objectif du jeu fait qu’un joueur a un intérêt fort à essayer de faire des propositions plutôt qu’à reprendre des phrases telles quelles, puisque ce dernier cas ne peut mener à l’attribution de points. Des éléments déterminants pour assurer la participation de joueurs incluent le fait que les parties soient courtes et facilement interruptibles, qu’elles portent sur des textes courts pour lesquels un réel besoin de réécriture existe, et que l’évaluation entre joueurs soit perçue comme ludique (en particulier, la définition de ce qui rend une paraphrase meilleure qu’une autre émergera du jugement des joueurs eux-mêmes). Une étude préliminaire a permis d’obtenir des jeux de reformulations pour des paragraphes issus de Wikipedia, qui montrent qu’avec des contraintes de réécriture assez libres, il est possible d’obtenir une grande variabilité dans les propositions soumises (cf. Figure 1). En outre, ce type d’interface peut également servir à des acquisitions ciblées pour des fragments de texte et/ou des contextes particuliers, si l’on propose des textes qui les contiennent.

3 Évaluation de paraphrases phrastiques et sous-phrastiques

Après une période spécifiée, il devient possible à des joueurs, qui n’ont pas nécessairement pris part à la première phase, de participer à l’évaluation d’une partie. Puisqu’un joueur doit juger plusieurs reformulations par phrase, il est important de limiter la quantité de phrases à lire et à évaluer en le dispensant d’avoir à formuler plusieurs fois un jugement sur des éléments communs. Un mode d’interaction dit de “graphe syntaxique” permet ainsi de visualiser les paraphrases à évaluer sous forme d’un graphe dirigé acyclique, comme illustré sur la Figure 1. Le graphe est obtenu par un algorithme de fusion/réduction guidé par la structure syntaxique inspiré de (Pang *et al.*, 2003). Un chemin complet dans le graphe correspond soit à la proposition d’un joueur, soit à une proposition inédite.

Dans ce mode, le joueur doit explicitement valider certains arcs (en gras sur la figure, les nœuds colorés indiquant des transitions dont tous les segments entrants et sortants ont été validés). La quantité de mots de la proposition d’un joueur appartenant à des arcs validés mène à un nombre de points proportionnel, avec une récompense particulière dans le cas où toute une phrase serait validée. Les reformulations acquises au niveau phrastique et sous-phrastique en contexte pourront obtenir des évaluations dépendant du nombre de jugements positifs rapportés au nombre de jugements total. En outre, ne seront pas pris en compte les jugements correspondant à des

¹Il est en outre possible de soumettre, après révision, le meilleur texte obtenu pour publication dans Wikipedia.

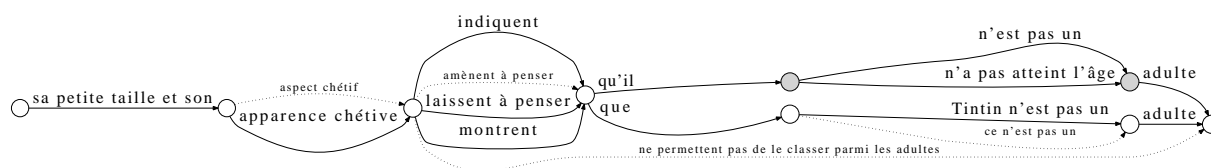


FIG. 1 – Exemple de “graphe syntaxique” pour la validation de 5 paraphrases pour la phrase : « *Ce n’est pas un adulte, comme le laisse à penser sa petite taille et son aspect chétif.* »

propositions faites par l’évaluateur lui-même, ainsi que les jugements de joueurs n’ayant pas acquis le statut de “joueurs certifiés”, statut accordé lorsqu’un joueur a montré des jugements suffisamment proches en moyenne de ceux des autres joueurs sur un certain nombre de parties. À l’issue de cette deuxième phase, les joueurs peuvent consulter les évaluations concernant leurs propres propositions et celles obtenant les meilleurs résultats, ainsi que confronter leurs jugements à ceux des autres évaluateurs. Différents classements (nombre de parties gagnées, nombre moyens de points par partie, moyenne des points des dix meilleures parties, nombres de phrases complètes retenues, etc.) sont accessibles et ont pour objectif de stimuler l’intérêt des joueurs et de les inciter à rejouer.

4 Discussion et conclusions

Les éléments clés du type d’application que nous avons présentée consiste en l’obtention d’une grande quantité de données associées à des jugements humains consensuels. Comme cela a déjà été proposé (Lafourcade, 2008), un jeu en ligne peut remplir ces objectifs. Cependant, l’application présentée permet également de mettre en place des campagnes d’acquisition de données ou d’évaluation de systèmes automatiques ciblées. Le jeu sera dans un premier temps proposé pour le français. Un des premiers usages que nous envisageons concerne l’acquisition de reformulations locales annotées comme étant des paraphrases ou non, qui seront utilisées comme une base d’exemples pour un apprentissage supervisé de classifieurs pour l’extraction de paraphrases depuis les révisions de Wikipedia, qui correspond à un travail en cours. Une autre utilisation concernera la réécriture automatique d’énoncés pour la traduction automatique et la recherche d’information par exploitation de paraphrases acquises depuis des sources hétérogènes. Enfin, les données produites seront régulièrement mises à disposition de la communauté afin d’aider la recherche sur les phénomènes paraphrastiques et leur évaluation.

Références

- COHN T., CALLISON-BURCH C. & LAPATA M. (2008). Constructing corpora for the development and evaluation of paraphrase systems. *Computational Linguistics*, **34**(4).
- LAFOURCADE M. (2008). JeuxDeMots : un prototype ludique pour l’émergence de relations entre termes. In *Actes de JADT*, Lyon, France.
- MAX A. (2008). Génération de reformulations locales par pivot pour l’aide à la révision. In *Proceedings of TALN*, Avignon, France.
- PANG B., KNIGHT K. & MARCU D. (2003). Syntax-based alignment of multiple translations : Extracting paraphrases and generating new sentences. In *Proceedings of NAACL/HLT*, Edmonton, Canada.