

# Sélectionner les “bons” passages pour créer les “bonnes” questions : Analyse et Évaluation d’un nouveau Corpus de Questions et Réponses pour l’Éducation

Thomas Gerald\*, Sofiane Ettayeb\*, Ha Quang Le \*\*, Gabriel Illouz\*\*\*, Patrick Paroubek\*\*\*, Anne Vilnat\*\*\*

\* Université Paris Saclay, CNRS, SATT Paris Saclay, LISN  
prenom.nom@lisn.fr

\*\* Professorbob.ai

ha-quang.le@polytechnique.edu

\*\*\* Université Paris Saclay, CNRS, LISN  
prenom.nom@lisn.fr

**Résumé.** Les systèmes intelligents pour le support scolaire sont aujourd’hui absents de la plupart des applications, alors que les récentes améliorations du Traitement Automatique des Langues (TAL) permettent d’imaginer des solutions innovantes. La création d’un système de questions-réponses reposant sur des sources scolaires permettrait d’accélérer, d’améliorer et de motiver l’apprentissage de l’étudiant. Dans ce contexte nous nous intéressons à la génération de questions au travers d’approches neuronales. Avec la récente création d’un corpus de questions-réponses par annotation de sources éducatives en langue française, nous disposons des ressources pour évaluer et développer de telles approches. Néanmoins, il faut considérer plusieurs obstacles : la quantité de données qualitatives n’est pas suffisante pour entraîner des approches génératives; dans le cadre d’une application autonome nous ne disposons pas explicitement du support pour la génération. Dans cette étude, nous proposons différentes méthodes d’extraction de ces supports comparant et analysant les résultats sur notre corpus et ceux de la littérature.

## 1 Introduction

Dans le milieu de l’enseignement, peu d’approches matures d’aide à l’apprentissage utilisant des algorithmes poussés d’apprentissage statistique sont aujourd’hui fonctionnelles. Les récentes avancées dans le traitement automatique des Langues (TAL) permettent pourtant de créer des outils pour traiter, extraire de l’information ou générer du contenu pour des documents et des supports de cours. Dans cet article, nous nous intéressons à cette problématique, en particulier en considérant les tâches de génération de questions et de création de réponses extractives ou abstractives. Nous travaillons en langue française afin de développer en partenariat avec l’industrie (entreprise ProfessorBob) un système automatique d’aide à l’enseignement. L’objectif poursuivi serait de fournir une application complète capable d’aider l’étudiant dans