

# Construction d'un corpus annoté en genre par apprentissage Zero-Shot.

Nicolas Béchet\*, Rémy Kessler\*\*, Gwen Icart\*\*\*, Gudrun Ledegen\*\*\*

\* UNIV. BRETAGNE-SUD, UMR CNRS 6074 IRISA,  
nicolas.bechet@irisa.fr

\*\* UNIV. BRETAGNE -SUD ,  
remy.kessler@univ-ubs.fr,,

\*\*\* UNIV. RENNES II, PREFICS  
gwenicart@outlook.com  
gudrun.ledegen@univ-rennes2.fr

**Résumé.** Afin de s'adapter au mieux au jeune public plus à l'aise avec les nouvelles technologies, une association a développé une application de webchat permettant à toute personne de partager ses angoisses. Plusieurs milliers de conversations anonymes ont ainsi été réunies et forment un corpus inédit de récits sur la détresse humaine, les violences sociales. Nous présentons dans cet article une méthodologie de production d'un modèle d'apprentissage permettant un étiquetage automatique en genre d'un corpus de texte en français. La méthode repose sur l'utilisation d'une combinaison d'un algorithme de classification Zero-Shot, d'une validation humaine et d'un apprentissage supervisé. Nous montrons que cette méthode permet de préannoter efficacement un corpus volumineux en présentant quelques résultats expérimentaux, validé par des experts.

## 1 Introduction

Depuis les années quatre-vingt-dix, la souffrance sociale est une thématique qui fait l'objet d'une grande attention de la part des pouvoirs publics ainsi que du milieu associatif. Parmi les conséquences, figure l'explosion des lieux d'écoute ou des dispositifs sociotechniques de communication dont les finalités consistent notamment à modérer les diverses formes de souffrance par la libération de la parole dans un but thérapeutique Fassin (2004, 2005). Une association de prévention du suicide a développé une application de webchat afin de répondre à ce besoin. Le webchat est un espace qui permet à toute personne d'exprimer et de partager avec un écoutant bénévole ses préoccupations, sa souffrance et ses angoisses. La principale spécificité de ce dispositif est son caractère non public et anonyme. Protégés par un pseudonyme, les appelants sont invités à confier auprès d'un bénévole les aspects problématiques de leur existence. Plusieurs milliers de conversations anonymes ont ainsi été réunies et forment un corpus inédit de récits sur la détresse humaine. Les travaux présentés dans ce papier s'inscrivent dans le cadre d'un projet de recherche lié à la prévention du suicide. Notre objectif dans ce projet est d'identifier différentes causes de souffrances parfois difficilement décelables, de caractériser ses modalités d'énonciation. Les premiers travaux réalisés ont notamment permis d'identifier