



UNIVERSITÉ  
DE LORRAINE



Master Informatique

# Rapport du projet d'initiation à la recherche

sujet : Constitution d'une base de cas de corrections du français

Année 2018-2019

Étudiants : Alex Ginestra  
Christopher Klein

Équipe : Orpailleur et  
Sémagramme  
Encadrants : Bruno Guillaume, Yves  
Lepage, Jean Lieber et  
Emmanuel Nauer

# Décharge de responsabilité

Nous soussignons Alex Ginestra et Christopher Klein, déclarons avoir pris connaissance de la charte des examens et notamment du paragraphe spécifique au plagiat.

Je suis pleinement conscient(e) que la copie intégrale sans citation ni référence de documents ou d'une partie de document publiés sous quelques formes que ce soit (ouvrages, publications, rapports d'étudiant, internet etc...) est un plagiat et constitue une violation des droits d'auteur ainsi qu'une fraude caractérisée.

En conséquence, je m'engage à citer toutes les sources que j'ai utilisées pour produire et écrire ce document.

# Remerciements

Tout d'abord, nous tenons à remercier toute l'équipe pédagogique du département informatique de la faculté des sciences et technologies pour les quatre années de formation en Master informatique.

J'adresse également mes remerciements à toute l'équipe Orpailleur et Sémaigramme ainsi qu'aux encadrants Bruno Guillaume, Yves Lepage, Jean Lieber et Emmanuel Nauer pour leur accueil et leur sympathie.

# 1 Sommaire

# Rappel du sujet

## Problématique de recherche :

Le raisonnement à partir de cas (RàPC) est un raisonnement hypothétique (en général) qui consiste à résoudre un nouveau problème (le problème cible, noté cible) en s'appuyant sur une base de cas, un cas étant la représentation d'un épisode de résolution de problème. On appelle cas source un élément de la base de cas. Souvent, on représente un cas source simplement par un couple (srce, sol(srce)) : srce est un problème source, sol(srce) est une solution de ce problème source. Un modèle du processus de RèPC classique comprend deux étapes d'inférence : Remémoration : un cas source (srce, sol(srce)) jugé similaire au problème cible (par exemple, sur la base d'une distance entre problèmes) est sélectionné. Adaptation : La solution sol(srce) de ce cas est modifiée en une solution candidate sol(cible) de cible.

La correction de phrases est la problématique de la transformation d'une phrase incorrecte (en particulier, grammaticalement) en une phrase corrigée (nous choisirons la langue française dans ce travail, même si la problématique existe dans toutes les langues). Un cas de correction de phrase est donc un couple (srce, sol(srce)) où srce est une phrase incorrecte et sol(srce) une correction de srce. Par exemple, on a les deux cas : srce1 = Tu as pas mangé. sol(srce1) = Tu n'as pas mangé. srce2 = Il a recommencer. sol(srce2) = Il a recommencé. L'adaptation se fait par des techniques de raisonnement par analogie : la solution sol(cible) est solution d'une équation analogique  $\hat{A} \ll \text{srce est à sol(srce)} \text{ ce que cible est à } y \hat{A}$ . Par exemple, l'adaptation de (srce2, sol(srce2)) à cible = Tu as manger. consiste à résoudre Il a recommencer. est à Il a recommencé. ce que Tu as manger. est à x qui a pour solution, avec la relation d'analogie utilisée dans le projet, x = Tu as mangé., proposition de solution proposée par le système. Sujet : Comme pour tout système à base de connaissances, la qualité d'un système de RèPC dépend de celle de son moteur d'inférences mais également de la qualité de sa base de connaissances, en particulier de sa base de cas. Une bonne base de cas doit avoir plusieurs qualités. Les cas sources doivent être corrects. Elle doit avoir une bonne couverture (et permettre de résoudre correctement une proportion importante de cas). Elle devrait ne pas être trop redondante (certains cas différents correspondent à la même correction). Pour cela, on pourra consulter les mouchards d'édition de Wikipédia pour en extraire des listes de fautes grammaticales ou orthographiques typiques, ainsi que des sites de dictées ou d'orthographe. Il faudra mettre en place les outils de collecte automatiques, paramétrables en fonction des sites. Une autre piste est la mise en place d'un jeu interactif avec un but. Le but est de faire corriger des phrases fautives par les joueurs. La phrase corrigée devrait émerger de la majorité des propositions de correction. Les phrases fautives pourraient être extraites de listes d'exemples fautifs, ou produites automatiquement à partir de patrons prédéfinis ou par application de l'analogie sur des cas déjà collectés. Une courte étude bibliographique sur la maintenance de base de cas permettra de suggérer des pistes pour l'acquisition d'une bonne base de cas. Il faudra mettre en place une méthode pour cela, qui pourra s'appuyer sur les sites mentionnés ci-dessus.

Bonjour, je lis actuellement un cours traitant de LaTeX!