

# French language corpus from TV shows

Project 1: A multilingual dialogue dataset. COMP-551: Applied Machine Learning, Fall 2017

Brian Choi  
ga.choi@mail.mcgill.ca  
McGILL ID:

Saif Rahman  
saif.m.rahman@mail.mcgill.ca  
McGILL ID:

Gerardo Tibamoso  
gerardo.tibamosopedraza@mail.mcgill.ca  
McGILL ID:

The corpus can be accessed at the following URL <http://mydataset.com>

## I. INTRODUCTION

We have gathered a collection of dialogues in French from scripts of the popular TV shows "Friends" (<http://www.fanfr.com/>) and "Extr@" (<http://channel4learning.com/>). The construction and main characteristics of this corpus are described as follows.

## II. DATASET DESCRIPTION

## III. DISCUSSION

## IV. STATEMENT OF CONTRIBUTIONS

We hereby state that all the work presented in this report is that of the authors.

## REFERENCES

- [1] I. V. Serban, R. Lowe, L. Charlin, and J. Pineau, "A survey of available corpora for building data-driven dialogue systems," *arXiv preprint arXiv:1512.05742v3*, 2017.