

Le grammairien: The  
Implementation of a  
System for Processing  
Object Clitics in French  
as a Second Language  
Based on AI

Adel Jebali

Concordia University, Montreal,  
Canada

# About the presenter

[adel.jebali@concordia.ca](mailto:adel.jebali@concordia.ca)

- ◆ Adel Jebali is an associate professor in Département d'études françaises at Concordia University (Montreal, Canada). With a Ph.D. in linguistics, he is interested in natural language processing, especially from an educational integration perspective.

# Outline



## Object clitics

French L2

Learners' difficulties

Learners' strategies



## Data

Sources

Composition

Examples



## Models

Model 1

Model 2

Grammairien

# Our goal

- ◆ To develop an automatic system (a computerized tool) capable of assisting learners in overcoming difficulties encountered in the production phase of object clitics in French as a second or foreign language (L2) by:
  - ◆ a) enabling them to distinguish correct productions from incorrect ones,
  - ◆ b) identifying the type of error made, and
  - ◆ c) reformulating the produced text containing the error by providing a correction.
- ◆ Part of a larger project: interlanguage processing

# State of the Art

- ◆ Interlanguage concept: Selinker (1972)
- ◆ Affes, Biskri and Jebali (2023): formalizing interlanguage in the ACCG framework
- ◆ Dodigovic (2003) and Maritxalar et al. (1997): focus on English and Basque
- ◆ Jebali (2023) and Parslow (2015): concentrate on French as a second language.
- ◆ Pienemann (1992): a multilingual approach aimed at second language acquisition researchers.

# Object Clitics in French L2

◇ Typical example:

Marie a mangé la pomme. (Mary ate the apple.)

→ Marie l'a mangée. (Mary ate it.)

l' (la) : object clitic, positioned before the auxiliary (different from the English counterpart it)

◇ The peculiar behavior of Object clitics in French, as well as other unique features of these elements → confusion among French L2 learners.

# Object Clitics in French L2

- ◇ Learners' difficulties:
  - ◇ Position: mostly preverbal, but postverbal in imperative
  - ◇ Agreement (in Gender, Number and Person) with the antecedent
  - ◇ Order (when more than one Object Clitic)
  - ◇ Semantics: human vs non-human objects (pets??)
  - ◇ Triggered agreement most of the time when preverbal
  - ◇ Etc.

# Object Clitics in French L2

- ◆ Learners' strategies:
  - ◆ Avoidance: Noun Phrase repetition at all costs
  - ◆ Null objects: Object Clitic omission
  - ◆ Paraphrase: *là-bas* instead of *y*, for example
  
- ◆ References : Jebali (2018), wust (2009)



# Data

- ◇ Need for 2 sets of data:
  - ◇ To train a model to be able to distinguish between correct and incorrect productions
  - ◇ To train a second one to be able to distinguish between different types of errors

# Data

## ◆ Sources

- ◆ Dataset1: primarily composed of authentic mistakes made by learners during a previous experiment conducted by the researcher (Jebali, 2018) + some synthetic productions that demonstrate common mistakes missing in our initial corpus.
  - ◆ Total of 2636 examples of various mistakes in the use of OCs; and
  - ◆ 2636 examples of perfectly correct productions sourced from various websites.
- ◆ Dataset2: productions containing four types of mistakes identified in the first dataset, supplemented with additional examples to create four balanced categories, each containing 1734 examples, for a total of 6936.

# AI Models

- ◇ Le grammairien: composed of 3 models
  - ◇ A fine-tuned FlauBERT (Le et al., 2020) model tailored for the task of grammaticality judgment and text classification → A binary classification of texts: correct/erroneous
  - ◇ A second model fine-tuned from the first one to make a four-types classification:
    - ◇ Type 1: Agreement errors
    - ◇ Type 2: Placement errors
    - ◇ Type 3: Strong pronouns vs clitics errors
    - ◇ Type 4: semantic and morphosyntactic errors

# AI Models

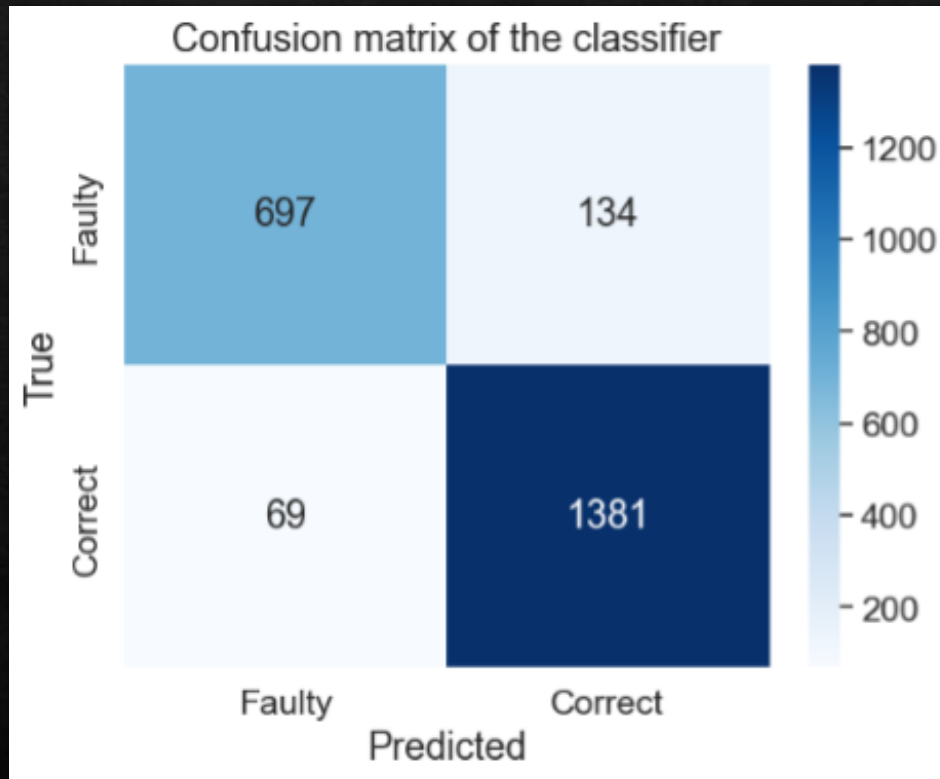
- ◇ The 3rd model: a generative one
  - ◇ A fine-tuned BARThez model (Kamal Eddine et al., 2021)
  - ◇ A new dataset in construction:
    - ◇ All the examples from the previous datasets
    - ◇ + examples from Lang-8, CEFLE and French Leiden Learner Corpus
  - ◇ At this time : the generative model
    - ◇ Can reformulate most of the erroneous productions regarding Object Clitics use
    - ◇ Can decently reformulate other errors as well, but more work on the dataset is needed

# The two AI models: comparison & evaluation

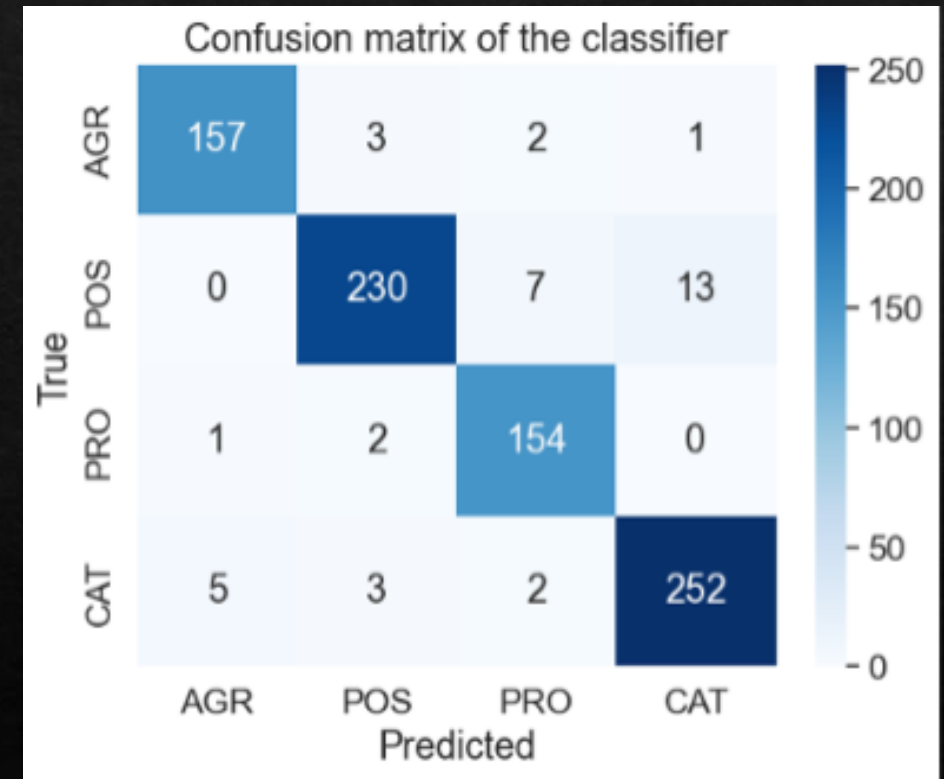
## 1. Comparison

	<b>Model1</b>	<b>Antidote 11</b>	<b>GPT 3.5</b>
Precision	.91	.84	.86
Recall	.84	.38	.54
F-score	.87	.52	.66

# 2. Evaluation



Model 1



Model 2

# Le Grammairien

## Le grammairien

Texte à corriger :

je suis aller voit Marie est j'ai lui aidée a emporte ses sac

« je » : féminin

Orthographe

Grammaire

Reformulation

Je suis allée voir Marie et je l'ai aidée à emporter ses sacs.

Copier

Log

# Key findings

- ◇ Object clitics in French L2 can be challenging for learners
- ◇ Avoidance and other strategies are adopted by them
- ◇ It is possible to handle a significant part of this hurdle by designing “le grammairien”:
  - ◇ It can handle different errors
  - ◇ It can correct most of them if need be