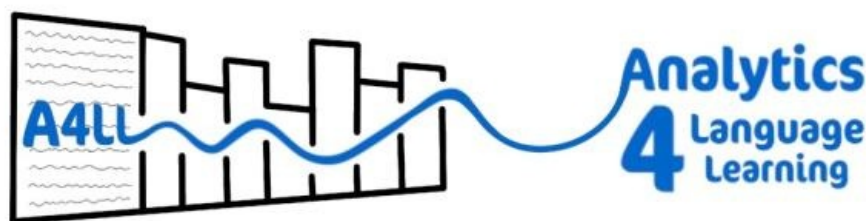
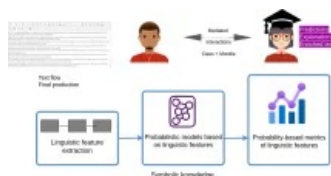


A4LL : Analytics for Language Learning



A4LL logo designed by Sidonie Tosser – Licence: CC-BY-NC 4.0

Why does my English teacher only ever underline the mistakes in my essays? Why does it take so long to correct my essay?



Description

The A4LL project will develop an innovative language-learning analytics system designed to assist teachers and learners with objective reports linking proficiency with linguistic features. Thomas Gaillat, the coordinator, proposes an approach relying on textual measures operationalising global and structure complexity, phraseology, discourse cohesion, and fluency. These measures will support the automatic creation of graphic reports used by teachers to diagnose their learners' productions. A4LL's ambition is to create the first fully automated L2 (Second language) analysis system serving learners, teachers, and researchers at university via an integrated data workflow from ingestion to analytics.

Research questions

The A4LL project will deliver an L2 analytics system for learners and teachers of English at university level. The project will address 3 main research questions aiming to uncover some of the features of Interlanguage, i.e the unstable linguistic system demonstrated by learners of a second language: i) what are the language features that are related to specific proficiency levels? ii) how can these features be measured automatically? iii) how can measures be converted into meaningful analytics for descriptive feedback and teaching decisions?

Interlanguage can be seen as a complex multifactorial system which makes the identification of criterial proficiency features difficult. Over time and practice, the system gradually stabilises. However, it is not clear which factors are at play at a given point. To cast a light on how interlanguage develops, current research shows that approaches combining linguistic measurements and statistics within computer models help to highlight some features of interlanguage (Ballier et al., 2020; Yannakoudakis et al., 2018). However, current state-of-the-art metrics lack linguistic meaningfulness and so impair interpretability.

Objective

The objective is to develop a computer system that automatically generates linguistic diagnostics of learner writings. These diagnostics will therefore be visualised by teachers through MOODLE, one of the main open-source LMS in France and in the world. These diagnostics will help teachers formulate advice for their students and adapt their teaching objectives in relation to their groups' profiles. Developing the system will imply research work to identify correlations between linguistic features and metadata including task types, proficiency, learning habits and writing ability. The system will collect, automatically analyse and provide specific linguistic feedback for writings submitted in MOODLE (see Figure 1). By exploiting lexical, syntactic and semantic metrics, the system will point out the dimensions that require attention in each writing. Graphical visualisations will show which linguistic areas to improve for a targeted proficiency level. The system will rely on a supervised learning approach with learner data collected in the two Language Centres (in charge of 20,000 students learning English for Specific Purposes) of the two universities of Rennes. It will be modular to allow subsequent integration of other languages.

A4LL intends to leverage the strength of two previously developed prototypes in which the coordinator participated. The first prototype, developed in 2019 (Gaillat, Simpkin, et al., 2021), provides [automatic classification of learner writings according to the levels of the CEFR](#). The second prototype, called [VizLing](#) (Gaillat, Knefati, et al., 2021), and developed in 2019, focused on the automatic generation of graphs to visualize linguistic complexity in writings. A4LL will expand in the same avenue, but it will rely on a selection of significant and linguistically descriptive metrics for second language analysis. A4LL will unify the Natural Language Processing tasks under a single framework producing visualisations in MOODLE. It will rely on learner metadata in order to allow teachers to profile their learners and personalise feedback.

The purpose of A4LL is thus i) to offer the language teaching community data analytics tools that help position learners according to proficiency and aspects of their language. ii) to model learner language to map linguistic features with proficiency and, ultimately, interlanguage stages. A4LL intends to provide a solution for university language centres, in France and abroad, that are in charge of millions of students studying languages for professional purposes.

Partners

Institution	Last Name	First Name	Role
Rennes 2 University	GAILLAT	Thomas	PI & Associate Professor
Rennes 2 university	MALLART	Cyriel	Research Engineer
Rennes 2 University	LI	Jen-Yu	Ph.D. candidate
Rennes 2 University	FAUGERE	Anatole	Research Assistant and Computer programmer
University of Paris Cité	BALLIER	Nicolas	Professor of Linguistics
University of Paris Cité	LISSON	Paula	Research Engineer
University of Galway	SIMPKIN	Andrew	Associate professors in Statistics
University of Galway	STEARNS	Bernardo	Research Associate
Le Mans University	VENANT	Rémi	Associate Professor
IRISA / INSA Rennes	SÉBILLOT	Pascale	Professor of Computer Science
IRISA / CNRS	GRAVIER	Guillaume	Senior Research Scientist

Partner project

Deep Learning for Language Assessment (DLLA)

Expert annotators

CEFR Annotation

Institution	Expert	Role	Structure
Rennes 2 University	Joanne Ward-Henry	English teacher	Centre de Langues Rennes 2
Rennes 2 University	Francoise Le Roux	English teacher	Centre de Langues Rennes 2
University of Rennes	Benedicte Dumont	English teacher	SCELVA, Univ de Rennes
University of Rennes	Pascale Janvier	English teacher	SCELVA, Univ de Rennes

Linguistic Annotation

- Team members: Paula, Nicolas and Thomas
- Université Paris Cité - CLILLAC-ARP: Jessica Tayeh

Conferences & publications

Deliverables

Software

- Gitlab LIDILE: The A4LL tool is available for download and installation (expertise required in Docker technology). It is made up of two components: [A4LL_MLpipeline and A4LL_Tool](#)
- A Tool for learner corpus collection: [The CELVA.SP User Interface for MOODLE](#)
- [The Google Colab](#) for the Workshop « [Enrichissement linguistique des données textuelles](#) », an Introduction to Python for Humanities Experts.

Supported by Rennes Métropole & ANR

Datasets & corpora

Learner corpus of language for Specific Purposes Three [datasets on Nakala](#):

- One with Dialang CEFR annotation
- Two batches with human expert CEFR annotation: 2018-2022 and 2023-2024

Credits: Many thanks to the language teachers of the universities of Rennes for their involvement
18 December 2024