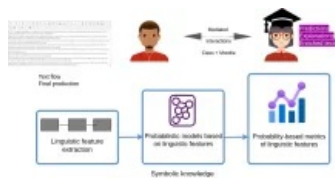




A4LL logo designed by Sidonie Tosser – Licence: CC-BY-NC 4.0

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	SÉBILLLOT	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

2025

Conference papers

titre

[Actionability in CALL: linking proficiency prediction models to interpretable indicators](#)

auteur

Thomas Gaillat, Cyrielle Mallart, Andrew Simpkin, Rémi Venant, Nicolas Ballier, Bernardo Stearns, Jen-Yu Li, Paula Lissón

article

*International Workshop on Foreign language learning and proficiency-rated reading materials: SLA research and AI methods supporting analysis and effective didactics in real-life education*, Universität Tübingen, Mar 2025, Tübingen, Allemagne, Germany

Accès au bibtex



titre

[L'usage des collocations en anglais d'apprenants : une analyse croisée des L1 et des niveaux de compétence](#)

auteur

Jen-Yu Li

article

*Approches interdisciplinaires des unités phraséologiques (UP) dans les langues du monde : Linguistique - TAL & IA - Traduction - Littérature*, Mar 2025, Paris, France

Accès au bibtex



Other publications

titre

[Annotated English Verb Noun collocation dataset](#)

auteur

Jen-Yu Li

article

2025

Accès au bibtex



titre

[CELVA.Sp processed with A4LL metrics pipeline](#)

auteur

Thomas Gaillat, Cyriel Mallart, Andrew J. Simpkin

article

2025, ([10.34847/nkl.3aba968r](#))

Accès au bibtex



Reports

titre

[Analytics for Language Learning Data Management Plan](#)

auteur

Thomas Gaillat, Nicolas Ballier, Cyrielle Mallart

article

Opidor. 2025, <https://dmp.opidor.fr/plans/13498>

Accès au bibtex



Preprints, Working Papers, ...

titre

[Assessing the validity of new paradigmatic complexity measures as criterial features for proficiency in L2 writings in English](#)

auteur

Cyriel Mallart, Andrew Simpkin, Nicolas Ballier, Paula Lissón, Rémi Venant, Jen-Yu Li, Bernardo Stearns, Thomas Gaillat

article

2025

Accès au texte intégral et bibtex



2024

Conference papers

titre

[La linguistique de corpus à l'heure du code ouvert](#)

auteur

Cyrielle Mallart, Thomas Gaillat, Rémi Venant, Nicolas Ballier, Jen-Yu Li, Bernardo Stearns

article

*Deuxième journée d'étude ARDoISE*, INRIA, Dec 2024, Rennes, France

Accès au bibtex



titre

[Evaluating the Generalisation of an Artificial Learner](#)

auteur

Bernardo Stearns, Nicolas Ballier, Thomas Gaillat, Andrew J. Simpkin, John P. Mc Crae

article

*NLP4CALL2024 : Natural Language Processing for Computer-assisted Language Learning*, Université Rennes 2, France; University of Gothenburg, Sweden; Linköping University, Sweden, Oct 2024, Rennes, France

Accès au texte intégral et bibtex



titre

[Overview of the linguistic features: creating measures – Joint presentation](#)

auteur

Nicolas Ballier, Bernardo Stearns, Jen-Yu Li

article

*pre-conference workshop to NLP4CALL 2024*, Oct 2024, Rennes, France

Accès au bibtex



titre

[Exploring learner knowledge with Large Language Models fine-tuned with the EFCAMDAT](#)

auteur

Nicolas Ballier, Bernardo Stearns

article

*LCR2024 Learner Corpus Research conference*, University of Tartu; Learner Corpus Association, Sep 2024, Tartu (Estonie), Estonia

Accès au texte intégral et bibtex



titre

[Assessing the validity of new structural complexity measures as features of proficiency in L2 English](#)

auteur

Thomas Gaillat, Cyriel Mallart, Nicolas Ballier, Andrew Simpkin, Rémi Venant, Bernardo Stearns, Paula Lissón, Jen-Yu Li

article

*Learner Corpus Research Conference*, University of Tartu, Sep 2024, Tartu (Estonie), Estonia

Accès au bibtex



titre

[Linguistic interoperability within a unified architecture](#)

auteur

Thomas Gaillat, Cyrielle Mallart, Andrew J. Simpkin, Rémi Venant, Nicolas Ballier, Jen-Yu Li, Bernardo Stearns

article

*Langues & Langage à la croisée des Disciplines - 1ère Rencontre annuelle LLcD*, Sorbonne Université; cnrs, Sep 2024, Paris, France

Accès au texte intégral et bibtex



titre

[Analytics for Language Learning. Linguistic interoperability within a unified architecture](#)

auteur

Cyriel Mallart, Andrew Simpkin, Rémi Venant, Nicolas Ballier, Bernardo Stearns, Jen-Yu Li, Thomas Gaillat

article

*Langues & Langage à la croisée des Disciplines 1ère Rencontre annuelle LLcD*, Sep 2024, Paris, France

Accès au bibtex



Other publications

titre

[Understanding Large Language Models](#)

auteur

Cyriel Mallart

article

2024

Accès au bibtex



titre

[Dictionary of Bigram-Score extracted from BNC with all association measures by NLTK](#)

auteur

Jen-Yu Li

article

2024

Accès au bibtex



Proceedings

titre

[Proceedings of the 13th Workshop on Natural Language Processing for Computer Assisted Language Learning](#)

auteur

Thomas Gaillat, Cyrielle Mallart, Fabienne Moreau, Griselda Drouet, Jen-Yu Li, David Alfter, Elena Volodina, Arne Jönsson

article

*The 13th Workshop on Natural Language Processing for Computer Assisted Language Learning*, Oct 2024, Rennes, France. LiU Electronic Press, 2024, Linköping electronic conference proceedings

Accès au texte intégral et bibtex



2023

Scientific blog post

titre

[CELVA.sp: A new learner language data set for the study of English for Specific Purposes at university level](#)

auteur

Thomas Gaillat, Cyrielle Mallart, Rémi Venant, Nicolas Ballier, Jen-Yu Li, Bernardo Stearns, Andrew Simpkin

article

2023

Accès au bibtex



Conference papers

titre

[L'interopérabilité des corpus pour la modélisation des dynamiques d'acquisition de langue seconde](#)

auteur

Thomas Gaillat, Cyrielle Mallart, Nicolas Ballier, Andrew Simpkin, Rémi Venant, Anatole Faugère, Bernardo Stearns, Jen-Yu Li, Paula Lissón

article

*Journée d'étude : « Corpus d'apprenants / corpus d'experts : Quels enseignements pour la caractérisation du discours scientifique ? »*, UR 3967 - CLILLAC-

Accès au bibtex



titre  
[Analytics for Language Learning: Interfacing MOODLE with A4LL via LTI](#)

auteur  
Thomas Gaillat, Cyrielle Mallart, Nicolas Ballier, Andrew Simpkin, Rémi Venant, Bernardo Stearns, Jen-Yu Li, Paula Lissón, Anatole Faugère

article  
*Deep learning for language assessment closing event (DLLA Closing event 2023)*, UR 3967 - CLILLAC-ARP : Centre de Linguistique Inter-langues, de Lexicologie, de Linguistique Anglaise et de Corpus-Atelier de Recherche sur la Parole; UFR EILA de l'Université Paris Cité, Nov 2023, Paris, France

Accès au bibtex



titre  
[Exploring a New Grammatico-functional Type of Measure as Part of a Language Learning Expert System](#)

auteur  
Cyriel Mallart, Andrew Simpkin, Rémi Venant, Nicolas Ballier, Bernardo Stearns, Jen-Yu Li, Thomas Gaillat

article  
*Proceedings of the 18th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2023)*, Jul 2023, Toronto, Canada. pp.466-476, ([10.18653/v1/2023.bea-1.39](#))

Accès au texte intégral et bibtex



titre  
[Analytics for Language Learning : Transmettre aux enseignants les profils linguistiques de leurs apprenants](#)

auteur  
Thomas Gaillat, Cyrielle Mallart, Anatole Faugère, Andrew Simpkin, Bernardo Stearns, Paula Lissón, Jen-Yu Li, Nicolas Ballier, Rémi Venant

article  
*Atelier GERAS @ 62e Congrès annuel de la SAES 2023*, Université Rennes 2; SAES La Sorbonne Nouvelle; GERAS (Groupe d'Etude et de Recherche en Anglais de Spécialité), Jun 2023, Rennes, France

Accès au bibtex



titre  
[Grammatical profiling with UD annotation \(WiP\)](#)

auteur  
Nicolas Ballier, Cyrielle Mallart, Thomas Gaillat

article  
*Workshop on Profiling second language vocabulary and grammar*, University of Gothenburg, Humanisten., Apr 2023, Gothenburg, Sweden

Accès au bibtex



Poster communications

titre  
[Exploring Verb-Noun collocations in learner English](#)

auteur  
Jen-Yu Li, Cyriel Mallart, Thomas Gaillat, Elisabeth Richard

article  
*Deep learning for language assessment (DLLA) closing event*, Nov 2023, Paris, France

Accès au texte intégral et bibtex



titre  
[Vers une grammaire probabiliste de micro-systèmes fonctionnels en L2](#)

auteur  
Cyrielle Mallart, Andrew Simpkin, Rémi Venant, Nicolas Ballier, Bernardo Stearns, Jen-Yu Li, Thomas Gaillat

article  
*RéAL2: Grammaire(s) et acquisition des L2: Approches, trajectoires, interfaces.*, Oct 2023, Grenoble, France

Accès au texte intégral et bibtex



2022

Conference papers

titre  
[Language learning analytics : designing and testing new functional complexity measures in L2 writings](#)

auteur  
Thomas Gaillat

article

*11th Workshop on Natural Language Processing for Computer-Assisted Language Learning (NLP4CALL 2022)*, Dec 2022, Louvain la Neuve, Belgium. pp.55-60, ([10.3384/ecp190006](https://doi.org/10.3384/ecp190006))

Accès au texte intégral et bibtex



2018

Software

titre

[CELVA.Sp corpus User Interface](#)

auteur

Thomas Gaillat, Rémi Venant, Cyriel Mallart, Taylor Arnold, Anatole Fougère

article

2018, ([swh:1:dir:7405005eae86eb3f53662e5649f10f5c4f92e11a;origin=https://gitlab.huma-num.fr/lidile/celva.sp-ui;visit=swh:1:snp:198c7b3333fa18b5a721d36e06e8a5a0648600e3;anchor=swh:1:rev:95c370947852a8fe6ef9254069ca7812fd901188](https://gitlab.huma-num.fr/lidile/celva.sp-ui;visit=swh:1:snp:198c7b3333fa18b5a721d36e06e8a5a0648600e3;anchor=swh:1:rev:95c370947852a8fe6ef9254069ca7812fd901188))

Accès au bibtex



## Deliverables

### Software

- 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