Research Day

Writing Practices in L2 Digital Ecosystems:

Japanese as a Field for Translinguistic Reflection

Programme

Organised by

Kaori OYA (IFRAE/UMR8043) and Jean BAZANTAY (PU, IFRAE/UMR8043)

As part of project :

Constitution and Exploitation of Language Corpora

(*Constitution et exploitation des corpus langagiers*)

Axe 1 : Languages, Sources, and the Issues at Stake
Theme 2 : Concepts and Words (*Les concepts et les mots*)

Of

IFRAE (French Research Institute on East Asia, UMR/8043)

Date : Wednesday, Septembre 9, 2026
Location : Auditorium Dumézil, Maison de la recherche
(Inalco, 2 rue de Lille – 75007 Paris)

9h30-9h35	**Opening Address** Jean Bazantay (IFRAE/UMR8043)
Session 1 : Exploration of W-CoLeJa Keystroke Logging Data (JP/EN)	
9h40 - 10h30 (50 min)	**Keynote address** **Writing Process Research in Japanese:** **Historical Perspectives and Future Directions** Professor Kei Ishiguro (National Institute for Japanese Language and Linguistics)
10h30 - 10h45 (15 min)	**Coffee Break**
10h45 - 11h20 (35 min)	**Kanji Input Revision Processes in Japanese Learners' Writing:** **A Log-Based Analysis** (Analysis of W-CoLeJa Keystroke Logging Data in Japanese) Yuzhe Li (Department of Japanese, University of Macau)
11h20-11h55 (35 min)	**The Production Process of Sentences in Learner Writing That Capture the Reader's Interest:** **An Analysis of the Writing Logs of One French-Speaking Learner** (Analysis of W-CoLeJa Keystroke Logging Data in Japanese) Yalan Zhu (National Institute for Japanese Language and Linguistics)
11h55 - 12h30 (35 min)	**Methodological Challenges in Keystroke Logging Analysis of L2 Japanese Writing:** **A Case Study of Pauses Preceding Nouns in French Learners of Japanese** (Analysis of W-CoLeJa Keystroke Logging Data in Japanese) Kaori Oya (Inalco/IFRAE)
12h30 - 14h00 (1h 30 min)	**Lunch**
Session 2: Exploration of keystroke logging data in other languages (EN)	
14h00 - 14h50 (50 min)	**Keynote address** **Linguistic Issues and Methodological Challenges in Analyzing the Writing Process** (Analysis of Keystroke Logging data in French) Pr Georgeta Cislaru (Sorbonne Nouvelle University / LLFL, ED 661, EA 7345)
14h50-15h30 (40 min)	**From Measurement to Feedback:** **Rethinking Writing-Process Research and Pedagogy in AI-Mediated Writing** Justin Young & Mariëlle Leijten (University of Antwerp)

15h30-15h40 (10 min)	**Coffee Break**
15h40-16h25 (45 min)	**Analysing Keystroke logging using the A4LL ecosystem** (Analysis of Keystroke Logging data in English) Pr Nicolas Ballier (Université Paris Cité / ALTAE), Ahood Al Sawar (Université Paris Cité / ALTAE), Thomas Gaillat (Université Rennes 2 / LIDILE), Rémi Venant (Université Le Mans / LIUM)
16h25-17h10 (45 min)	**Data analysis and visualisation in keystroke logging research** **Online workshop** Marina Beccard (Chemnitz University of Technology) Christina Sanchez-Stockhammer (Chemnitz University of Technology)
17h10-17h45 (35 min)	**Using keystroke logs to inform feedback on the writing process** Léo Nebel (Sorbonne University / LIP6)
17h45-18h05 (20 min)	**Coffee Break**
	Summary
18h05-18h25 (20 min)	**Round Table:** **Didactic Challenges of Contemporary Keyboard-Assisted Writing Practices in the Digital Age** Moderation: Jean Bazantay (Inalco/IFRAE)
18h25-18h30 (5 min)	**Concluding Words**

Writing Practices in L2 Digital Ecosystems:
Japanese as a Field for Translinguistic Reflection

Jean Bazantay (PU, IFRAE/UMR8043)

In the teaching of Japanese as a foreign language in France, the development of writing skills has traditionally relied on students composing short sentences, followed by progressively longer texts. Students' written work is generally assessed with a strong emphasis on grammatical accuracy and character formation. While feedback may occasionally highlight the relevance of the content and the stylistic quality of certain passages, guidance tends to remain ad hoc rather than being integrated into a sustained, longitudinal approach to the development of writing skills. This approach rests on the implicit assumption that writing competence — understood both as the acquisition of the writing system and as the development of discursive competence — develops primarily through practice. The processes underlying the development of these competences are rarely made explicit, let alone systematically examined. The emphasis is instead placed on repeated practice, with writing tasks gradually increasing in length and complexity, but with little sustained reflection on the cognitive and discursive processes involved.

Furthermore, although theoretical models of L1 and L2 writing have incorporated the transformations brought about by digital technologies and keyboard-mediated writing (Hayes, 1996, 2012; Chenoweth & Hayes, 2001; Leijten & Van Waes, 2014), their effects on the different subprocesses involved in writing remain to be fully investigated. In the case of Japanese, keyboard-mediated writing involves specific competences, such as mastery of input and conversion operations, including the use of predictive input tools (Matsumoto-Stuart, 2003; Iwashita & Nakahara, 2019; Oya, forthcoming). Moreover, in digital environments, learners draw on external resources in the process of composing their texts, which has significant implications for the acquisition of writing competence. While some researchers have begun to explore the use of such resources (Knospe, 2017) and the associated language choices (Kruse, 2024) using keystroke logging software (Leijten & Van Waes, 2013), research on L2 writing in Japanese has so far relied primarily on questionnaires, stimulated recall interviews, or screen recordings (Aikawa, 2018; Tanaka, 2015; Suzuki, 2012, 2014, 2020). It is precisely these questions that this research day aims to address by bringing together researchers who explore new methodological approaches to investigating how learners construct their texts and make use of the digital tools available to them, from both acquisitional and didactic perspectives.

One emerging line of research draws on data collected during the construction of learner corpora, such as the W-CoLeJa corpus developed by Ishiguro (2026) at the National Institute for Japanese Language and Linguistics. Such data provide valuable resources for overcoming some of the limitations of oral elicitation procedures, whether concurrent or retrospective. Among these data, log files generated by keystroke logging software — which record the complete history of keyboard input, keystroke by keystroke — make it possible to reconstruct the writing process as it unfolds over time. These files allow researchers to track errors,

revisions, pauses, and writing bursts, thereby providing valuable insights into the cognitive dynamics of writing. However, due to their size and complexity, the analysis of such data requires appropriate methodologies and remains an area of ongoing development.

To mark the publication of the W-CoLeJa corpus, this research day aims to explore the potential of log file data for research by bringing together analytical approaches and methodologies developed across different research teams. The event seeks to combine scientific analysis, pedagogical reflection, and interdisciplinary perspectives, with particular attention to how process data can contribute to a fine-grained understanding of foreign language writing activity within digital ecosystems.

The presentations in the morning session, entitled “Exploration of W-CoLeJa Keystroke Logging Data” (JP/EN), will focus on research conducted using data collected during the construction of the W-CoLeJa learner corpus. The session will open with a presentation of the corpus by Professor Ishiguro, who led the project at NINJAL. The following presentations will examine specific cases of process data analysis based on the log files collected during the construction of the corpus. The first study will address issues related to conversion operations in non-Latin writing systems. It will highlight some of the specific features of keyboard input in languages using different writing systems, an issue that is relevant to many of the languages taught at Inalco. The second study will focus on the qualitative dimension of the final product, particularly its ability to engage the reader, examining this dimension in relation to the text construction process as reconstructed from log files. The final presentation will focus on the temporal analysis of the writing process and, in particular, on the location of pauses, especially before nouns. Particular attention will be paid to their interpretation as possible indicators of cognitive effort or planning activity, depending on the context in which they occur.

Across these presentations, issues of segmentation and the granularity of units of analysis — character, morpheme, word, phrase — will be investigated, alongside patterns of correction and revision, whether local or global, immediate or delayed. Japanese thus provides a basis for broader translinguistic reflection on writing processes in digital environments.

The afternoon session will broaden the perspective by presenting research conducted in different European contexts and involving different first and foreign languages. Sharing a common interest with research on Japanese in the analysis of writing-process data, these studies will provide points of comparison and methodological insights that may inform research on Japanese as well as on other languages.

The first presentation, by Georgeta Cislaru, will provide an overview of the work conducted within the ANR Pro-TEXT project, which seeks to model the textualization process through the analysis of writing activity recorded in real time. Particular attention will be paid to the relationship between pauses and linguistic units, as well as to the identification of units of written performance. The presentation will also address several methodological issues involved in processing such data — pause segmentation, preprocessing, annotation, and visualization — and their potential applications to the teaching of writing in both first and foreign languages.

The following presentation will introduce the A4LL (Analytics for Language Learning) ecosystem, developed at the University of Rennes, which enables the longitudinal collection of learner texts and keystroke data. Particular attention will be paid to the processing pipeline for log files — conversion, annotation, and the automatic extraction of metrics — as well as to the potential of these data for analysing writing behaviour and modelling learner proficiency. Already implemented for English, French, and Swedish, the system also opens up possibilities for other languages, including Japanese. Justin Young's presentation will focus on the transformations brought about by generative AI in digital writing practices. Drawing on data collected with Inputlog, it will examine both the methodological challenges that AI-integrated writing poses for writing-process research and the potential pedagogical uses of such data. In particular, the presentation will show how process data can help learners become more aware of their own writing strategies and develop more deliberate approaches to AI-mediated writing.

An online workshop will then focus on methods for the statistical analysis and visualization of keystroke logging data, ranging from the selection of appropriate statistical tests and units of analysis to different ways of representing writing processes and revision behaviour. It will thus provide practical methodological tools for analysing and interpreting process data. Finally, the last presentation, "Using Keystroke Logs to Inform Feedback on the Writing Process," will extend this pedagogical perspective by examining how keystroke data can be used to provide learners with feedback on their writing processes.

Taken as a whole, this second session will bring together different approaches to collecting, processing, analysing, and using writing-process data, highlighting their relevance both to research on writing and to the teaching of writing.

The research day will conclude with a round table on "Didactic Challenges of Contemporary Keyboard-Assisted Writing Practices in the Digital Age."

The event is organised as part of the project Building and Using Language Corpora (Axis 1: Languages, Sources and Their Stakes; Theme 2: Concepts and Words) at the Institut Français de Recherche sur l'Asie de l'Est (UMR 8043).

References:

Bazantay, J. (2024). Constitution d'un corpus longitudinal de rédactions d'apprenants de japonais. *Actes du 18e colloque sur l'enseignement du japonais en France organisé par l'Université Grenoble-Alpes les 14 et 15 juin 2024 (AEJF Symposium Ronshū 3)*. AEJF.

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication, 18 (1)*, 80–98.

Cislaru, G., & Olive, T. (2018). *Le processus de textualisation : Analyse des unités linguistiques de la performance écrite*. De Boeck Supérieur.

Gunnarsson-Largy, C. (2022). *La production écrite en français L2 : Focus sur le processus de formulation* [Doctoral dissertation, Université Toulouse – Jean Jaurès]. HAL. tel-03640707.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 1–27). Erlbaum.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*, 369–388. DOI

Ishiguro, K., Sano, A., & Ji, T. (2023). Dictionary tools used by learners of Japanese around the world: Towards appropriate support of vocabulary search behavior on smartphones. *The Japanese Journal of Language in Society, 26*(1), 5–20.

Ishiguro, K., & Wūri Zhé. (2025). *Gakushūsha kōpasu no sekkei to kōchiku* [Conception et constitution d'un corpus d'apprenants]. Kenkyūsha.

Knospe, Y. (2017). *Writing in a third language: A study of upper secondary students' texts, writing processes and metacognition* [Doctoral dissertation, Umeå University]. Umeå Studies in Language and Literature, 36.

Kruse, M. (2024). Problem-solving activity during the foreign language writing process: A proposal for categorisation and visualisation of source use and a new take on fluency in multilingual writing. *Journal of Writing Research, 16*(1), 129–161. DOI

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research: Using Inputlog to analyze and visualize writing processes. *Written Communication, 30*(3), 358–392.

Leijten, M., & Van Waes, L. (2020). Designing keystroke logging research in writing studies. *Chinese Journal of Second Language Writing, 1*(1), 18–39.

Leijten, M., Van Waes, L., Schriver, K., & Hayes, J. R. (2014). Writing in the workplace: Constructing documents using multiple digital sources. *Journal of Writing Research, 5*(3), 285–337. DOI

Matsumoto-Stuart, Y. (2003). Nihongo gakushūsha ni yoru wāpro bunsho no goyō kanji wa "dōon kanji no gohenkan" na no ka: Hi-kanjiken gakushūsha no goyō hyōki bunseki [Are misused kanji in Japanese learners' word-processor texts homophone mis-conversions? An analysis of errors among non-kanji learners]. *Japanese Language Education, 118*, 17–26.

Nakahara, K., & Iwashita, M. (2019). Nihongo gakushūsha ni okeru Nihongo kanji meishi sanshutsuji no shutsuryoku keishiki no eikyō: Kanjiken gakushūsha to hi-kanjiken gakushūsha no hikaku [Effects of output modality on the production of Japanese kanji nouns: A comparison of learners from kanji and non-kanji backgrounds]. *Bulletin of Nagasaki University of Foreign Studies, 23*, 43–51.

Oya, K. (forthcoming). Le processus d'écriture médiatisée par un clavier en japonais. In *Japon Pluriel 16 : Actes du 16e colloque de la Société Française des Études Japonaises*. Éditions Picquier.

Wang, Z., & Han, F. (2017). Metacognitive knowledge and metacognitive control of writing strategy between high- and low-performing Chinese EFL writers. *Theory and Practice in Language Studies, 7*(7), 523–532.

Wang, W., & Wen, Q. (2002). L1 use in the L2 composing process: An exploratory study of 16 Chinese EFL writers. *Journal of Second Language Writing, 11*(3), 225–246.

Session 1 :

Exploration of W-CoLeJa Keystroke Logging Data (JP/EN)

「パソコンによる L2の文章産出過程
―外国語としての日本語教育（JFL）環境で学ぶ学習者を対象に―」

講演

「日本語における文章産出過程研究―研究史の振り返りと将来の展望―」

石黒 圭（国立国語研究所）

Writing Process Research in Japanese: Historical Perspectives and Future Directions

ISHIGURO Kei (National Institute for Japanese Language and Linguistics)

本講演は、日本語における文章産出過程の研究の歴史と現在、そして将来の展望について広く議論することを目的としている。

日本語の文章産出過程研究の歴史には一定の蓄積があり、また、欧米とは異なる独自の発展を遂げてきた点に特徴がある。研究に着手した当初は学習者の作文執筆が主な対象であり、分析方法も発話プロトコル法を用いた内観法が主流であったが、その後、母語話者の執筆行動に調査対象が広がり、分析方法も手書きやタイピングといった、より客観的に記録できる外部記録法へと移ってきた。

また、日本語の文章産出過程研究の現在を考える場合、本研究チームが分析に用いている EssayLogger に触れる必要がある。日本語版キーストロークロギングのシステム開発は、ある学習者作文コーパス構築のための調査のさい、アプリ開発者の新城直樹氏（琉球大学）が副次的に実装した簡素な執筆記録システムがきっかけとなった。そのログデータを見て、日本語学習者の文章産出過程の可能性に気づいた講演者が研究を進めると同時に、世界的な日本語学習者作文コーパスである W-CoLeJa を構築するさい、当初のシステムの精度を高めた EssayLogger を執筆システムに組み込み、執筆の速度、修正、休止の様子を記録できるようにした。その結果、修正の種類や休止の長さ、修正や休止の位置や理由を考慮した総合的な分析が可能になり、多くの研究者によって活用されている。

将来の展望としては、日本語教育のみならず、GIGA スクール構想によってタブレット普及が進む日本の国語教育の現場においてもキーストロークログを収集し、第一言語における文章産出過程の縦断的発達過程を解明することが期待される。一方、研究面では、執筆時間も考慮に入れた文章産出過程の定量的・総合的な分析への道が開かれていると考える。

Writing Process Research in Japanese: Historical Perspectives and Future Directions

ISHIGURO Kei (National Institute for Japanese Language and Linguistics)

This lecture aims to provide a broad discussion of the history, current state, and future directions of research on writing processes in Japanese.

Research on writing processes in Japanese has accumulated a substantial body of scholarship and has developed along a trajectory distinct from that of research conducted in Europe and North America. In its early stages, the field primarily focused on composition writing by learners of Japanese, with introspective approaches based on think-aloud protocols serving as the principal method of analysis. Subsequently, the scope of research expanded to include the writing behavior of native Japanese speakers, while analytical methods gradually shifted toward external recording techniques that allowed writing activity, including handwriting and typing, to be documented more objectively.

Any discussion of the current state of Japanese writing process research must also consider EssayLogger, the tool used by our research team. The development of a Japanese-language keystroke-logging system originated in a simple writing-recording program created by Naoki Shinjo of the University of the Ryukyus as a supplementary component of a learner corpus project. After examining the resulting log data, the speaker recognized its potential for investigating the writing processes of Japanese language learners and began developing this line of research. During the construction of W-CoLeJa, a global longitudinal corpus of Japanese learner writing, an enhanced version of the original system, named EssayLogger, was incorporated into the writing platform. This system makes it possible to record writing speed, revisions, and pauses. As a result, researchers can conduct comprehensive analyses that take into account the types and duration of revisions and pauses, as well as their locations and possible causes. EssayLogger has consequently been adopted by a growing number of researchers.

Looking ahead, keystroke logs are expected to be collected not only in Japanese language education but also in first-language Japanese education, where the use of tablets has become increasingly widespread under Japan’s GIGA School Program. Such data may help clarify the longitudinal development of first-language writing processes. From a research perspective, these developments are also opening the way for quantitative and comprehensive analyses of writing processes that incorporate writing time as an important variable.

「日本語学習者の作文における漢字入力の修正過程―ログデータにもとづく分析―」

李 羽喆（マカオ大学）

本研究は、キーストローク・ログを用いて、日本語学習者の作文における漢字入力の修正過程を明らかにすることを目的とする。分析の中心となるのは、「不採用漢字語」、すなわち、いったん入力されたものの、確定後に削除されたり、再変換によって別の語に置き換えられたり、言い換えによって削除されたりしたため、最終的な文章には残らなかった漢字語である。これまで多くの研究では誤用や最終的な産出物が分析されてきたが、こうした不採用となった形式に注目することで、最終的な文章からは見えてこない、学習者のリアルタイムの語彙選択や迷いを捉えることができる。

分析には、縦断的な日本語学習者作文コーパスである W-CoLeJa core_05 のデータを用いる。このコーパスには六つの言語圏の学習者が含まれており、漢字圏・非漢字圏の双方の学習者を対象としている。ログデータには、入力の系列、変換候補の選択、再変換、削除、置換、および時間情報が記録されている。修正が生じた位置を再現可能な形で分類するため、本研究では形態素にもとづく基準を採用する。すなわち、入力中の形態素の内部で行われた変更を「入力レベルの修正」、いったん次の形態素へ進んだ後、前の箇所に戻って行われた変更を「文内修正」として分類する。

そのうえで、それぞれの不採用漢字語を最終的に選択された表現と対応づけ、両者について語彙レベル、意味領域、文体の観点から比較する。さらに、第一言語（L1）の背景および日本語習熟度によるパターンの違いを比較するとともに、処理負荷を示す指標として、再変換の回数や修正に要した時間などの時間的指標についても検討する。

予想される結果としては、たとえば、より難易度の高い漢語からより平易な表現への変更といった、一定の方向性をもつ語彙置換が確認される可能性がある。また、第一言語の異なる学習者群によって、迷いの現れ方にも異なる特徴が見られることが予想される。本研究の成果により、学習者がデジタル環境で作文を書く際に、漢字変換をどのように処理し、語彙選択をどのように修正しているのかが明らかになるとともに、日本語教育におけるフィードバックの設計や作文支援ツールの開発にも示唆を与えることが期待される。

Kanji Input Revision Processes in Japanese Learners' Writing: A Log-Based Analysis

Yuzhe Li（University of Macau）

This study examines kanji input revision processes in Japanese learners' writing using keystroke logs. The focus is on "unadopted kanji words," that is, kanji words that were typed but did not remain in the final text because they were deleted after confirmation, replaced during reconversion, or removed through rephrasing. While many studies describe errors and final products, these unadopted forms show learners' real-time lexical decisions and uncertainty that are not visible in the final text.

The data comes from the W-CoLeJa core_05 longitudinal learner writing corpus, which includes writers from six linguistic regions and covers both kanji background and non-kanji background learners. The logs record input sequences, conversion candidate selections, reconversion, deletions, replacements, and timing information. To code revision location in a replicable way, we adopt a morpheme-based rule: changes made within an ongoing morpheme are coded as input level revision, while changes made after the writer has moved to the next morpheme and then returns are coded as within sentence revision. We then link each unadopted kanji word to the final chosen expression and compare their lexical level, meaning domain, and style. We also compare patterns by L1 background and proficiency level, and examine timing measures such as reconversion counts and revision duration as indicators of processing load.

Expected findings include systematic replacement directions, such as shifts from higher level Sino Japanese items to simpler forms, and different profiles of hesitation across L1 groups. The results will clarify how learners manage kanji conversion and revise lexical choices in digital writing, and will inform feedback design and writing support tools for Japanese language education.

「心を惹かれた学習者作文の文の産出過程
―フランス語話者の執筆ログ分析を通して―」

朱 雅蘭（国立国語研究所）

本発表では、「第二言語として日本語を学ぶ学習者は、どのように日本語を用いて、読み手の心を惹く作文を書いているのか」という問いを出発点とする。

発表者は、2024 年 10 月より、『日本語学習者縦断作文コーパス』（W-CoLeJa）の構築作業の一環として、中国語話者やフランス語話者などによる日本語作文の形態素解析作業に携わっている。その作業を進める中で、フランス語話者の作文に触れる機会が多くあり、その中には「なぜか頭から離れない」と感じる作文がしばしばあった。

本発表では、こうしたフランス語話者の作文をはじめ、発表者自身が読みごたえがあると感じた作文を「心を惹かれた作文」と定義する。そして、中国人日本語学習者であり、文章論を志す大学院生でもある発表者自身の視点から、フランスの大学で日本語を学ぶ学習者1名による「心を惹かれた作文」のログデータを対象とし、その作文が完成するまでの執筆過程において、学習者がどのような試行錯誤を重ねているのかを質的に記述する。

予想される結果として、石黒・朱（2026）で取り上げた中国語話者の「心を惹かれた作文」の執筆過程と比較すると、フランス語話者には、「挿入」によって論述を展開・補完していく傾向がみられると考えられる。この「挿入」を中心とした執筆過程を追跡することで、フランス語の作文教育で用いられる「ディセルタシオン」という弁証法的な文章構造が、第二言語としての日本語作文の執筆過程にも反映されている可能性を検討する。

本発表の意義は、発表者自身の文化的背景とはやや距離のあるフランス語話者による「心を惹かれた作文」の執筆過程を質的に記述することを通して、フランス語話者ならではの日本語作文の執筆過程の特徴と魅力を提示する点にある。

参考文献

石黒圭・朱雅蘭(2026)「日本語学習者の文の産出過程の追跡―ある学部生の執筆過程ログを手がかりに―」『政大日本研究』23, 1-38.

The Production Process of Sentence in Learner Writing That Captures the Reader's Interest: An Analysis of the Writing Logs of One French-Speaking Learner

ZHU Yalan (National Institute for Japanese Language and Linguistics)

This study begins with the following question: How do learners of Japanese as a second language use Japanese to produce compositions that capture the reader's interest?

Since October 2024, the presenter has been involved in the morphological analysis of Japanese compositions written by Chinese-speaking, French-speaking, and other learners as part of the development of the *Longitudinal Corpus of Japanese as a Second Language* (W-CoLeJa). Through this work, the presenter has encountered a considerable number of compositions written by French-speaking learners, some of which have remained particularly memorable and difficult to forget.

In this study, compositions by French-speaking learners, as well as other texts that the presenter personally found engaging and rewarding to read, are defined as "learner compositions that capture the reader's interest." From the perspective of the presenter, who is both a Chinese learner of Japanese and a graduate student specializing in writing studies, the study qualitatively describes the writing process of one such composition produced by a learner studying Japanese at a French university. Using the learner's writing-log data, the analysis examines the forms of trial and error through which the composition was gradually produced.

In comparison with the writing process of a Chinese-speaking learner's composition examined by Ishiguro and Zhu (2026), the French-speaking learner is expected to demonstrate a tendency to develop and supplement the discussion through the insertion of additional material. By tracing this insertion-centered writing process, the study explores the possibility that the dialectical structure of the *dissertation*, a form commonly used in French writing education, may also be reflected in the learner's process of composing in Japanese as a second language.

The significance of this study lies in its qualitative description of the writing process underlying a composition by a French-speaking learner whose cultural background differs to some extent from that of the presenter. In doing so, it seeks to identify and demonstrate the distinctive characteristics and appeal of the Japanese writing processes of French-speaking learners.

Reference

Ishiguro, K., & Zhu, Y. (2026). Tracing the sentence production processes of a learner of Japanese: An analysis based on the writing-process logs of an undergraduate student. *Chengchi Journal of Japanese Studies, 23*, 1–38.

Methodological Challenges in Keystroke Logging Analysis of L2 Japanese Writing:

A Case Study of Pauses Preceding Nouns in French Learners of Japanese

Kaori OYA (INALCO/IFRAE)

This presentation aims to illustrate the methodological challenges involved in analysing keystroke logging data collected using EssayLogger (Inoue, 2026) to investigate the development of language-specific processing in L2 Japanese writing. It explores how pauses and P-bursts can be measured and analysed in L2 Japanese writing while taking into account the use of external resources, drawing on data collected over two academic years from two French learners of Japanese at different proficiency levels, with the aim of applying this analytical approach to a larger sample.

Examining second language (L2) learners' pausing behavior provides important insights into the cognitive processes involved in writing (Barkaoui, 2019). L2 writers tend to pause more frequently at lower textual units (Spelman Miller, 2000; Leijten & Van Waes, 2015), reflecting language-related processing (Révész et al., 2026). Accordingly, L2 writers tend to produce shorter P-bursts, which become longer with greater linguistic experience (Lindgren et al., 2008). In digital writing, pauses may also be associated with the use of external resources. However, it remains unclear how pauses and bursts can be measured and analyzed in Japanese writing, whose transcription system introduces additional cognitive demands, such as selecting the appropriate character type and conversion candidate in digital writing (Oya, 2025). This issue may be particularly relevant for learners whose L1 is French, a language that differs from Japanese in terms of both linguistic structure and transcription system.

Building on a preliminary analysis (Oya, forthcoming), which showed that pauses at lower units occurred most frequently before noun production, this presentation focuses on pauses preceding nouns and on the content within P-bursts while also taking into account pauses associated with the use of external resources such as online dictionaries for noun production.

Keystroke logging data and screen recordings were collected as part of the W-CoLeJa longitudinal learner writing corpus at INALCO. A 2-second threshold, based on previous L2 writing research, was used as an analytical starting point to extract pauses preceding nouns. Sentence-internal morphological boundaries were identified using Web ChaMame with Mecab (version 0.996) and UniDic for Contemporary Written Japanese (version 202512), and the content within P-bursts was manually annotated. Screen recordings were annotated using ELAN to identify and visualize when external resources were consulted.

Preliminary analysis revealed changes in the number of pauses and the content within P-bursts, which may reflect an expansion in the range of language-specific processing that learners can perform without pausing, from lexical retrieval and character transcription to conversion candidate selection and syntactic encoding. This expansion also appeared to vary with learners' proficiency levels. The use of external resources for noun production was not limited to the period immediately preceding nouns. After consulting these resources, the

learners also engaged in syntactic processing based on search results, with individual variation in both resource use and subsequent processing.

References:

Barkaoui, K. (2019). What can L2 writers' pausing behavior tell us about their L2 writing processes? Studies in Second Language Acquisition, 41, 529–554.

Inoue, Y. (2025). 入力インターフェイスの開発　Development of an input interface. In Ishiguro, K. & Wu, R (Eds.). 学習者コーパスの設計と構築 Design and construction of a learner corpus. Kenkyûsha, 64-76.

Leijten, M. & Van Waes, L. (2015). Fluency in Writing: A Multidimensional Perspective on Writing Fluency Applied to L1 and L2. Computers & Composition, 38, 79-95.

Lindgren, E., Sullivan, K.P.H. & Spelman Miller, K. (2008). Development of fluency and revision in L1 and L2 writing in Swedish high school year 8 and 9. International Journal of Applied Linguistics, 156, 133-151.

Oya, K. (2025). Le processus d'écriture médié par un clavier en japonais. Le 16ème colloque de la Société Française des Études Japonaises.

Oya, K. (forthcoming). 日本語学習者の文章産出過程における執筆休止の縦断的変化―フランス語話者の執筆ログ分析から― Longitudinal Changes in Pausing During the Writing Process of Japanese Language Learners: A Keystroke Logging Data Analysis of French-Speaking Learners. [Manuscript submitted for publication], In Yokono, H., Asai, T., & Ishiguro, K. (Eds.), Writing Process and Assessment. Kenkyusha.

Révész, A., Jeong, H., Liu, Y., Otsuki, T., Rong, W., Zeng, T., Takahashi, M., Sugiura, M. (2026). The neural correlates of pausing behaviours in L2 writing. Journal of Seconde Language Writing, 72, 1-13.

Spelman Miller, K. (2000). Academic writers on-line: Investigating pausing in the production of text. Language Teaching Research, 4, 123-148.

Web ChaMame (MeCab version 0.996; UniDic for Contemporary Written Japanese, version 202512). National Institute for Japanese Language and Linguistics. https://chamame.ninjal.ac.jp/

Session 2 :

Exploration of keystroke logging data in other languages (EN)

Linguistic Issues and Methodological Challenges in Analyzing the Writing Process

Georgeta Cislaru, Clesthia, Université Sorbonne Nouvelle

This presentation will provide an overview of the work conducted as part of the ANR Pro-TEXT project, which aims to model the process of textualization through the analysis of the writing process recorded in real time (Cislaru & Olive 2018). While the project lies at the intersection of linguistics, psycholinguistics, and computer science/NLP, the central issue is primarily linguistic, as it seeks to elucidate the linguistic mechanisms involved in text production. However, the various behavioral and temporal phenomena specific to the process, as well as the temporal, spatial, and linguistic constraints, present significant methodological challenges.

We will first address the linguistic issues involved in the study of the writing process. Since we are dealing with texts of varying lengths, we are faced with a discursive flow that – we know from studies of spoken language – cannot be perfectly fluid and contains pauses or disfluencies. The question, therefore, is whether the alignment keylogging data such as pauses and linguistic units is significant in terms of the nature of these units or in terms of their role in segmenting units of written performance (Cislaru et al. 2025). These questions, which fall within the realm of theoretical linguistics, are fully applicable in practice. On the one hand, the identification of units of written performance can be utilized in the teaching of writing, both in the native language and in a foreign language (Feltgen & Lefeuvre 2025, Métairy & Cislaru 2026). On the other hand, examining the sentence as a unit of performance helps to highlight syntactic (and not just orthographic) difficulties in children with dyslexia (Cislaru et al. 2024).

However, the nature of the results obtained depends on several factors, such as the pause segmentation method (Bouriga & Olive 2021, Feltgen & Cislaru 2024), the preprocessing of recorded data, the ability to annotate the data, and visualization resources (Cislaru et al. 2026). Each of these challenges will be discussed in detail and related to the issues in linguistic analysis.

References

Bouriga, S. et Olive, T. 2021. Is Typewriting More Resources-Demanding than Handwriting in Undergraduate Students ? *Reading and Writing* 34 9Q : 2227-2255. DOI : https://doi.org/01007/s11145-021-10137-6

Cislaru, G., Doquet, C., Olive, T., 2025, Operating Units in Written Language Performance: Linguistic and Behavioral Perspectives. *Journal of Writing Research*, *16*(3)*,* 359-375.

Cislaru, G., Feltgen, Q., Khoury, E., Delorme, R., Bucci, M.-P., 2024, "Language processing units are not equivalent to sentences: evidence from writing tasks in typical and dyslexic children", *Languages* Languages | Free Full-Text | Language Processing Units Are Not Equivalent to Sentences: Evidence from Writing Tasks in Typical and Dyslexic Children (mdpi.com)

Cislaru, G., Grozavu, N., Olivié, M., 2026, « Towards a 3-sided behavioral/linguistic/Machine Learning Analysis and Visualization of key-log data », in Christophe Leblay, Hakim A. Usoof, Gilles Caporossi (éds) An Introduction to Data Visualisation of the Writing Process, Leiden, Brill, Studies in Writing 41, p. 126-149. doi:10.1163/9789004760264_007

Cislaru, G., Olive, T. 2018, *Le processus de textualisation.* Bruxelles : De Boeck.

Feltgen, Q., Cislaru, G., 2024, « The fluency vs. disfluency dichotomy in writing processes as reflected in the structure of the inter-key intervals empirical distribution », *Discourse Processes* 62 (1), Taylor & Francis, p. 16-39.

Feltgen, Q. et Lefeuvre, F. 2025. Clitic Subjects as Landmarks in the Writing Production Process : A Study Based on a Keylog-Derived Corpus of Writing Bursts. *Journal of Writing Research* 16 3Q : 435-462. DOI : https://doi.org/017239/jowr-2025.603.04

Métairy, J., Cislaru, G., 2026, Les blocs constructionnels comme unités de traitement à l'écrit : le cas de la construction attributive du sujet, *Discours* 37. DOI : https://doi.org/10.4000/15t57 <https://journals.openedition.org/discours/14066>

From Measurement to Feedback:

Rethinking Writing-Process Research and Pedagogy in AI-Mediated Writing

Justin Young & Mariëlle Leijten, The University of Antwerp

Abstract:

Generative AI is reshaping digital writing in ways that challenge established writing-process research methods while creating opportunities for process-oriented pedagogy. This presentation examines both issues through the discussion of Inputlog-based research and a classroom workshop on AI-mediated writing. It examines the methodological challenges AI-integrated writing presents for writing process researchers, particularly when GenAI complicates the interpretation of source events, pauses, fluency, revisions, and keystroke activity by functioning not only as a digital source, but also as a tool for generating, transforming, evaluating, and reorganizing text. The presentation then shows how these data can be repurposed pedagogically. In a Digital Communication course, multilingual Master students in professional communication completed an AI-integrated writing task in English and reflected on their approach. They then received instruction on effective AI-mediated writing practices derived from the analysis of Inputlog writing-process data from an expert AI user. The presentation will describe how Inputlog-supported reflection helped make students' own writing strategies visible and gave them a basis for comparing their processes with expert practice and developing more deliberate, transferable approaches to AI-mediated writing.

Analysing Keystroke logging using the A4LL ecosystem

Ahood Al Swar, Université Paris Cité / ALTAE, ahood-saleh-m.al-sawar@etu.u-paris.fr

Nicolas Ballier, Université Paris Cité / ALTAE, nicolas.ballier@u-paris.fr

Thomas Gaillat, Université Rennes 2 / LIDILE, thomas.gaillat@univ-rennes2.fr

Rémi Venant, Université Le Mans / LIUM

This talk will present the A4LL ecosystems and its moodle-based infrastructure designed to collect learner essays, metadata and keylogs using moodle. We will present how data was collected for English learners over three years at the University of Rennes over three years using the A4LL infrastructure. This Analytics for Language Learning (A4LL) system was conceived during an ANR-funded research project and aims at providing a dashboard for students and teachers alike. For the time being, the dashboard reports statistical modelling of CFER levels of students based on lexical and syntactic complexity, but keylogging data is captured with the hope of fine-tuning the statistical models on the basis of keylogging.

In the A4LL infrastructure, we use a separate free JSON-based keylogger that does not capture any specific user IDs and preserves learner anonymity (and is therefore compliant with Europe's GDPR requirements for data collection). The JSON-based system has proved to be robust and has been used for a keylogging corpus (Velentzas et al, 2024). We have developed a pipeline to convert the JSON-based keylogging files into easier .CSV datasets using the tidy data format as explained in (Pacquetet, 2024). We have devised a certain number of automatic metrics designed to analyze learner production and ultimately to contribute to learner-level evaluation. We will demo the infrastructure and show that it can be applied to the analysis of several languages, English, French, Swedish (as already implemented at the University of Rennes), but potentially for Japanese as well.

After these longitudinal data collection examples, we will present some of the metrics we have implemented in our infrastructure to analyse learner keylogging data. We will explain our pre-processing phase (from .JSON format to .CSV), our annotation phase (how we have tried to incorporate a Universal Dependency annotation phase to keylogging data) and our post-processing phase, where we automatically extract metrics to be used for exploratory data analysis and, in the future, for learner modeling (Al Swar et al, 2025). In line with previous research on keylogging modelling (Conjin et al, 2019a, 2019b), we have incorporated a copy task and a writing task to our data collection pipeline. We will illustrate some of our metrics and discuss the relevance of behavioural metrics (Tappert et al, 2010).

References:

Al Swar, A., Pacquetet, E., Mallart, C., Simpkin, A. J., & Ballier, N. (2025). Predicting CEFR levels for learners of English with keylogging metrics, an exploratory study. In *CORIA-TALN 2025* (pp.7-11) ⟨hal-05146826⟩.

Conijn, R., Roeser, J., & Van Zaanen, M. (2019a). Understanding the keystroke log: The effect of writing task on keystroke features. *Reading and Writing*, *32*(9), 2353-2374.

Conijn, R., Van Zaanen, M., Leijten, M., & Van Waes, L. (2019b). How to typo? Building a process-based model of typographic error revisions. *Journal of Writing Analytics*, *3*, 69-95.

Pacquetet, E. 2024. *The effect of linguistic properties on typing behaviors and production processes.* Ph.D. thesis, University at Buffalo.

Tappert, C. C., Villani, M., & Cha, S. H. (2010). Keystroke biometric identification and authentication on long-text input. In *Behavioral biometrics for human identification: Intelligent applications* (pp. 342-367). IGI Global Scientific Publishing.

Velentzas, G., Caines, A., Borgo, R., Pacquetet, E., Hamilton, C., Arnold, T., Ballier, N. & Yannakoudakis, H. (2024). Logging keystrokes in writing by English learners. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)* (pp. 10725-10746).

Online Workshop

Data analysis and visualisation in keystroke logging research

Marina Beccard & Christina Sanchez-Stockhammer (Chemnitz University of Technology)

Overview

This workshop introduces participants to key approaches for analysing and visualising keystroke logging data in Japanese as a Second Language writing research. The session is divided into two parts. The first focuses on statistical methods commonly used in keystroke logging studies, while the second explores effective strategies for visualising writing-process data and research findings.

Data analysis

The first part of the workshop provides an overview of statistical procedures frequently employed in keystroke logging research. Particular attention will be paid to selecting appropriate statistical tests and understanding their underlying assumptions.

Topics include:

- Determining sample size before data collection through a priori power analysis.
- Descriptive statistics, including means and standard deviations, as well as raw and relative frequencies (percentages).
- Inferential statistics, with a focus on tests such as ANOVA and ANCOVA (e.g., as in Al-Saadi and Galbraith (2021))
- Assessing inter-rater reliability when multiple researchers code or review revision events, including the use of Krippendorff's alpha (Conijn et al., 2022).

Visualisation

The second part of the workshop addresses the visual representation of keystroke logging data. Participants will discuss how to select appropriate visualisations and apply fundamental design principles to create figures that are clear, informative, and accessible.

Topics include:

- Dot plots with and without regression curves (e.g., Al-Saadi & Galbraith, 2021).
- Process graphs generated using Inputlog (e.g., Al-Saadi & Galbraith, 2021).
- Bar / stacked bar charts
- Visualisations using S-notation to represent revision behaviour (Chan & Lam, 2025).

Overall, the workshop seeks to support researchers in making informed methodological choices and in communicating findings from keystroke logging studies in a clear and compelling manner.

References

Al-Saadi, Z. T., & Galbraith, D. (2021). Does revision process differ across language of writing (L1 vs. FL), FL language proficiency and gender? Writing & Pedagogy, 12(1), 73-109.

Chan, S., & Lam, D. M. (2025). Zooming in on the process-product nexus of meaning-related revisions: A micro-analytic approach to keystroke logging data. Journal of Second Language Writing, 69, 101228.

Conijn, R., Speltz, E. D., Zaanen, M. V., Waes, L. V., & Chukharev-Hudilainen, E. (2022). A product-and process-oriented tagset for revisions in writing. Written Communication, 39(1), 97-128.

Using keystroke logs to inform feedback on the writing process

Léo Nebel, Sorbonne Université, LIP6, leo.nebel@lip6.fr

July 20, 2026

1 Abstract

Many studies in Automated Writing Evaluation examine feedback on writing tasks, but they often limit their scope to the final written product. Hattie and Timperley emphasize that, for feedback to be effective, the process level is crucial, alongside the task level, the self-regulation level, and the self level [1]. Automatically offering guidance on the writing process is, however, particularly difficult: students differ widely in their needs, and there is rarely a single, universally valid set of writing strategies. Previous research has investigated how keystroke logging can support feedback on the writing task by taking the process into account. Vandermeulen et al. [4] primarily aimed to "mirror" students' own writing processes back to them through visualizations and indicators. Their feedback was supplemented with questions and comparisons to the processes of other students in order to prompt reflection on writing strategies. Building on this work, and relying on an open-source Python library for characterizing revision processes[1], we present our recent study [2], which extends this mirroring approach by testing new visualizations and a new type of guiding feedback. We report on an experiment with 184 10th-grade students in French (L1) classes. We found correlations between students' progress in writing performance and the feedback condition (students who received process-focused feedback performed slightly better than those who did not), and we observed clear differences in how their writing strategies evolved. While we will briefly discuss these findings, the presentation will also address the open challenges involved in providing feedback on the writing process. We argue that keystroke logging offers substantial potential for generating more adaptive and timely [3] feedback. Regarding adaptivity, a promising direction for future work is the development of a writing-specific learner model that directly informs feedback. Concerning timeliness, we envision using keystroke logging features that can be computed on the fly to deliver feedback during the writing activity itself.

[1] https://pypi.org/project/krev/

References

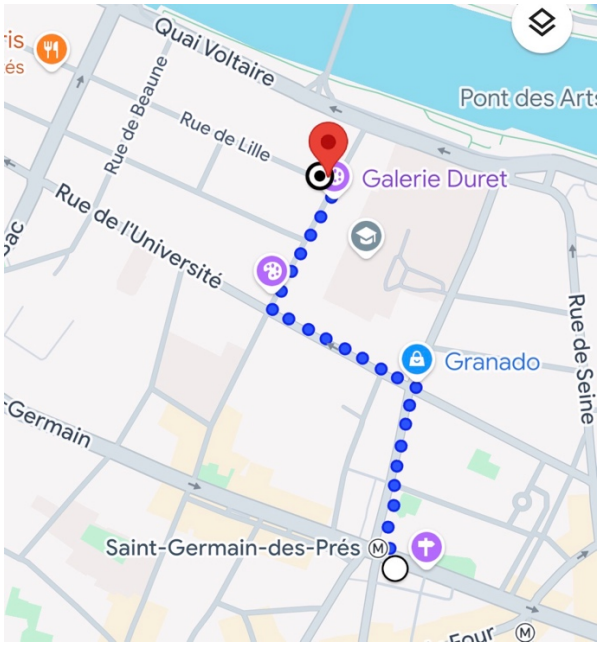
[1] J. Hattie and H. Timperley. The power of feedback. *Review of Educational Research*, 77(1):81–112, 2007.

[2] L. Nebel, F. Bouchet, and V. Luengo. From mirroring to guiding: The impact of process feedback on revision strategies. In E. G. Blanchard, G. Chen, M. Chi, and S. Isotani, editors, *Artificial Intelligence in Education. Late Breaking Results, WideAIED, Practitioners, Industry and Policies, Blue Sky, Doctoral Consortium, FoL Workshops and Tutorials, FoL Invited Papers*, pages 79–85, Cham, 2026. Springer Nature Switzerland.

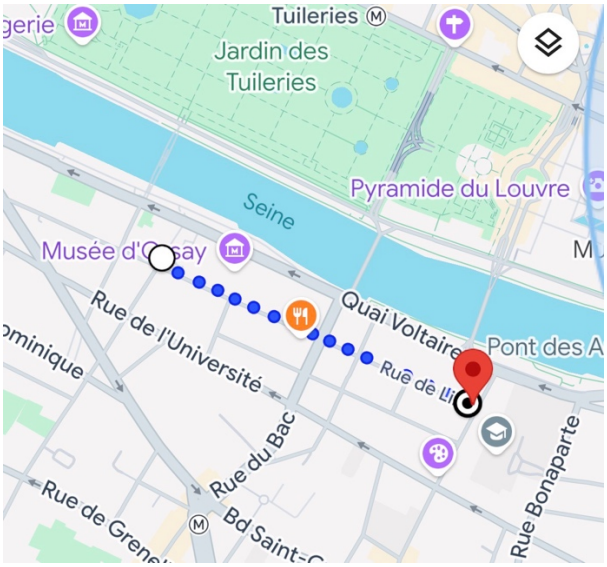
[3] V. J. Shute. Focus on formative feedback. *Review of Educational Research*, 78(1):153–189, 2008.

[4] N. Vandermeulen, M. Leijten, and L. Van Waes. Reporting writing process feedback in the classroom: Using keystroke logging data to reflect on writing processes. *Journal of Writing Research*, 12(1):109–140, 2020.

Auditorium Dumézil, Maison de la recherche
(Inalco, 2 rue de Lille – 75007 Paris)



8-minute walk from **Métro station Saint-Germain-des-Prés**



9-minute walk from **RER C station Musée d'Orsay**