

Vers une analyse temporelle des processus rédactionnels des élèves

Léo Nebel¹[0000–0001–5859–265X], François Bouchet²[0000–0001–9436–1250] et Denis Alamargot³[0000–0001–9807–3723]

¹ EvidenceB leo-n@evidenceb.com

² Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
francois.bouchet@lip6.fr

³ Université de Poitiers - INSPE de l'académie de Poitiers Laboratoire CHArt-UPEC
denis.alamargot@univ-poitiers.fr

Résumé. L'évaluation formative des productions écrites dans un cadre scolaire est un sujet complexe. Cet exercice mobilise chez les élèves de nombreuses compétences, dont notamment la maîtrise des processus rédactionnels. De nombreux outils existent déjà pour aider les professeurs dans leurs évaluations, voire dans certains cas pour automatiser l'évaluation d'un texte. Cependant, il n'existe pas d'outil à notre connaissance qui exploite des traces temporelles afin d'analyser la maîtrise des processus rédactionnels par les élèves, en parallèle des méthodes d'évaluation statiques plus classiques. Nous examinons ici les apports potentiels d'une telle analyse temporelle dans le cadre d'une évaluation formative d'écrits d'élèves.

Mots-clé: traitement automatique de la langue naturelle · processus rédactionnels · analyse temporelle · révision · évaluation formative

Abstract. The formative evaluation of written work in a school setting is a complex subject. This exercise mobilizes many skills for students, including the mastery of writing processes. Many tools already exist to help teachers in their evaluations, even in some cases to automate the evaluation of a text. However, to our knowledge, there are no tools that exploit time traces to analyze students' mastery of writing processes, in parallel with more traditional static assessment methods. We examine here the potential contributions of such a temporal analysis in the context of a formative assessment of students' writing.

Keywords: natural language processing · writing processes · temporal analysis · revising process · formative assessment

1 Introduction

L'évaluation représente une grande partie du temps du professeur. À travers le développement des usages d'outils numériques au sein de milieux éducatifs,

son automatisa-tion devient de plus en plus courante. Cette automatisa-tion est toutefois plus compliquée dans le cadre d'exercices incluant du texte libre. Des modèles basés sur des données statistiques existent depuis longtemps, d'autres plus récents se fondent aussi sur des méthodes de *deep learning* [8]. Les corrections proposées par ces modèles peuvent porter sur le style du texte, sa forme, son sens et/ou sa pertinence. Ces corrections correspondent à celles réalisées par le professeur qui n'a accès qu'à la production finale d'un élève. Mais au-delà du gain de temps apporté par les outils numériques, ils permettent de suivre le déroulé de l'activité de rédaction de l'élève rendant possible une analyse automatique. On peut alors se poser les questions de recherche suivantes : (1) cette analyse automatique dynamique permet-elle d'enrichir la qualité de l'évaluation ? (2) permet-elle d'améliorer aussi les rétroactions fournies par l'EIAH (Environnement Informatique pour l'Apprentissage Humain), notamment les rétroactions formatives et motivationnelles ?

2 État de l'art de l'évaluation de texte d'élèves

De nombreux travaux ont déjà été réalisés dans le champ de l'évaluation automatique d'essais (AES pour *Automated Essay Scoring*, au caractère sommatif) et d'écrits (AWE pour *Automated Writing Evaluation*, au caractère formatif) [7].

Les premiers systèmes d'évaluation se concentraient sur des critères liés au style de la production écrite. Ils s'intéressaient notamment à la diversité du vocabulaire à travers des indicateurs comme le ratio *Type-Token*, la correction orthographique ou la complexité syntaxique [3]. Ces évaluations de forme sont rendues possibles depuis un certain temps grâce à différentes techniques de traitement automatique du langage naturel (tokenisation, analyse morpho-syntaxique, dictionnaires...). Des systèmes plus récents évaluent aussi le fond et la sémantique des productions. La vectorisation des mots, des phrases et des textes à l'aide de modèles pré-entraînés basés sur l'architecture Transformers tels que BERT [5] ou d'autres modèles similaires [4] ont permis ces nouveaux résultats.

Peu de systèmes mêlent ces deux modes d'évaluation (fond et forme). Si les modèles de *deep learning* prennent indirectement en compte la forme dans leur notation (des corrélations entre forme et fond existant dans un nombre important de productions), cette évaluation statique ne permet pas de comprendre la stratégie rédactionnelle de l'élève et l'enchaînement des processus dans le temps.

Il est ainsi possible d'ajouter une troisième dimension à cette évaluation : l'analyse temporelle. La plupart des systèmes d'évaluation se limitent à évaluer l'état final du texte (le produit). Or dans un objectif d'évaluation formative, l'évaluation de la maîtrise du discours et de la gestion temporelle est essentielle [2]. De nombreux travaux ont montré que les bons écrivains mobilisaient et enchaînaient des processus de rédaction spécifiques [1]. À travers les outils pédagogiques numériques, nous pouvons mettre en place des systèmes retraçant de manière précise l'évolution temporelle d'un texte. À notre connaissance, peu de modèles ajoutent cette composante dans leur évaluation. Même s'il existe des recherches où des feedbacks ont été donnés à partir d'analyses temporelles [6], ces

dernières ne proposent pas de modèles automatisés permettant cette évaluation ni une évaluation croisée avec les critères plus classiques discutés ci-dessus.

3 Contexte et modélisation

Nous proposons ainsi de tester une évaluation dynamique afin d’observer les avantages que cette dernière pourrait apporter dans la notation et le feedback écrit proposé aux élèves. Cette évaluation aura un double caractère dynamique : (1) à travers l’analyse de l’évolution du texte mais aussi (2) à travers le suivi continu de l’élève et la mesure de ses progrès, cette deuxième partie n’étant pas traitée ici. Nous nous appuyons sur un modèle d’exercice dans lequel on donne à l’élève une introduction et une conclusion, et où on lui demande de faire le lien entre les deux en rédigeant un texte. Cet exercice est réalisé par des élèves de lycée dans un exercice plus général visant la remise à niveau sur différentes compétences. L’outil développé évalue le texte rédigé selon 6 critères : (1) la sémantique, (2) la syntaxe, (3) l’orthographe, (4) le vocabulaire, (5) la fluence, (6) la révision. Les quatre premiers critères s’évaluent traditionnellement uniquement à partir de l’état final (statique) du texte tandis que les deux derniers nécessitent de suivre son évolution (dynamique). Cette évaluation de la dynamique nécessite une collecte des frappes claviers, des temps associés et de la position du curseur.

Concernant la fluence, l’enjeu est d’évaluer la vitesse d’écriture de l’élève comparée à sa vitesse d’écriture “habituelle”. Cela peut se faire à partir des données collectées et d’un comptage dynamique des mots écrits via une tokenisation. Pour le critère de révision, l’enjeu consiste à construire un score cohérent à partir de multiples indicateurs pour quantifier la tendance à la révision de l’élève. Parmi ces indicateurs on trouve le nombre et le temps médian de pause, le ratio de mots produits par rapport aux mots processés (c’est-à-dire les mots qui ont été effacés par la suite), ou encore la distance sémantique entre les différents états du texte après chaque phase de relecture. Le score de révision se calcule par une moyenne pondérée de l’ensemble des indicateurs. Cette pondération se fait de manière équilibrée mais sera sujette à un calibrage à l’avenir. À partir de ce score, un retour global sera donné à l’élève, ainsi que des retours plus précis sur chacun des indicateurs afin de ne pas perdre en information.

Au début de notre modélisation, nous avons collecté des données auprès d’un public varié, constitué d’une cinquantaine de réponses, accompagnées des traces clavier permettant d’effectuer nos analyses dynamiques. Ces données nous ont permis d’établir une première version d’un système expert qui a pour prochain objectif d’être ajusté auprès d’un public cible d’élèves plus important.

4 Prochaines étapes

L’étape immédiate à venir sera de valider la première version de notre modèle. Une expérimentation va être réalisée dans ce cadre auprès d’un public cible d’élèves. La progression des élèves à travers plusieurs sessions sera mesurée et

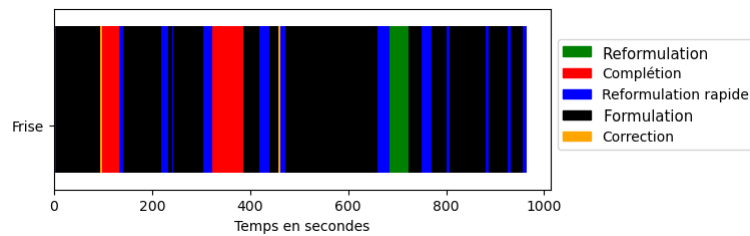


Fig. 1. Représentation temporelle des transitions d'états du texte d'un élève

comparée selon les différents cas envisagés. On pourra ainsi évaluer l'influence de l'analyse dynamique en complément de l'analyse statique par rapport à une simple analyse statique.

À moyen terme, il serait intéressant de construire à partir de nos différents indicateurs, des représentations temporelles de l'évolution du texte (cf. Figure 1). Ce découpage permettrait de mieux rendre compte des stratégies rédactionnelles existantes et de leur influence sur la qualité générale du texte. Dans un second temps, une collecte massive de traces pourra être effectuée à travers une application éducative. Ces données, une fois étiquetées, nous permettront de construire automatiquement des représentations temporelles des réponses d'élèves potentiellement plus précises que celles construites à partir de nos simples indicateurs. Des expérimentations permettront de confirmer cette hypothèse.

References

1. Alamargot, D., Lebrave, J.L.: The Study of Professional Writing. *European Psychologist* **15**(1), 12–22 (Jan 2010). <https://doi.org/10.1027/1016-9040/a000001>
2. Alamargot, D., Plane, S., Lambert, E., Chesnet, D.: Using eye and pen movements to trace the development of writing expertise: case studies of a 7th, 9th and 12th grader, graduate student, and professional writer. *Reading and Writing* **23**(7), 853–888 (Aug 2010). <https://doi.org/10.1007/s11145-009-9191-9>
3. Blache, P.: Un modèle de caractérisation de la complexité syntaxique. *traitement automatique des langues naturelles* pp. 1–10 (2010)
4. Chang Lu, M.C.: Integrating deep learning into an automated feedback generation system for automated essay scoring. *Proceedings of The 14th International Conference on Educational Data Mining (EDM 2021)* pp. 573–579 (2021)
5. Devlin Jacob, Chang Ming-Wei, L.K., Kristina, T.: Bert: Pre-training of deep bidirectional transformers for language understanding. pp. 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota (2019)
6. Nina Vandermeulen, Elke Van Steendam, S.D.M., Rijlaarsdam, G.: Writing process feedback based on keystroke logging and comparison with exemplars: Effects on the quality and process of synthesis texts. *Written communication* **40**, 90–144 (2023)
7. Ramesh Dadi, S.K.S.: An automated essay scoring systems: a systematic literature review. *Artificial Intelligence Review* **55**, 2495–2527 (2022)
8. Uto, M.: A review of deep-neural automated essay scoring models. *Behaviormetrika* **48**, 459–484 (2021)