



HAL
open science

Analyser automatiquement les signaux de l'enseignement : Une approche d'apprentissage social fondée sur les preuves

Romain Laurent, Philippe Dessus, Dominique Vaufreydaz

► To cite this version:

Romain Laurent, Philippe Dessus, Dominique Vaufreydaz. Analyser automatiquement les signaux de l'enseignement : Une approche d'apprentissage social fondée sur les preuves. A.N.A.E. Approche neuropsychologique des apprentissages chez l'enfant, 2022, 176, pp.29-36. hal-03599280v1

HAL Id: hal-03599280

<https://hal.univ-grenoble-alpes.fr/hal-03599280v1>

Submitted on 7 Mar 2022 (v1), last revised 13 Mar 2022 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**Analyser automatiquement les signaux de l'enseignement :
Une approche d'apprentissage social fondée sur les preuves**

Romain Laurent^{1,2}, Philippe Dessus^{1,*}, & Dominique Vaufreydaz²

1. Univ. Grenoble Alpes, LaRAC, 38000 Grenoble, France

2. Univ. Grenoble Alpes, LIG, CNRS, Grenoble INP, 38000 Grenoble, France

* Auteur correspondant :

Philippe Dessus

Bât. Michel-Dubois

CS 40700

38058 Grenoble CEDEX 9

France

philippe.dessus(@)univ-grenoble-alpes.fr

Adresse du 1^{er} auteur :

Romain Laurent

LaRAC

Bât. Michel-Dubois

CS 40700

38058 Grenoble CEDEX 9

France

romain.laurent(@)univ-grenoble-alpes.fr

Analyser automatiquement les signaux de l'enseignement : une approche d'apprentissage social fondée sur les preuves

Résumé : Les récentes avancées en traitement et analyse du signal ont permis de créer de nouvelles manières d'instrumenter l'observation et l'analyse des événements scolaires, et donc de recueillir de nouveaux types de preuves des pratiques d'enseignement ou d'apprentissage. Cet article recense certains en se fondant sur le cadre d'analyse de l'apprentissage social, posant que la pédagogie est une activité sociale intégrée à la vie de tous les jours, et reposant sur certaines capacités humaines innées.

Mots-clés : Apprentissage social ; Éducation fondée sur les preuves ; Pédagogie ; Traitement et analyse du signal ; Apprentissage machine

Automated Analysis of Instructional Signals: An Evidence-based Social Learning Approach

Abstract: Recent advances in signal processing and analysis have made it possible to create new ways of instrumenting the observation and the analysis of educational events, and thus to gather new kinds of evidence on teaching and learning practice. This article identifies some of these, based on a “social learning” framework, which posits that pedagogy is a social activity embedded in everyday life, and relies on certain innate human capacities.

Key-Words: Social Learning; Evidence-based Education; Pedagogy; Signal Processing and Analysis; Machine Learning

Introduction

Cet article propose une nouvelle manière de considérer et de capturer des preuves issues de la pratique d'enseignement. Toute personne (chercheur, enseignant, formateur, inspecteur) intéressée à caractériser cette pratique et à recueillir des données à son propos dispose de moyens qui sont limités à l'observation en situation de cours (directe ou enregistrée), au recueil d'opinions *via* questionnaires et à l'observation participante (de type recherche-action). Ces moyens ont tous des inconvénients : l'observation nécessite la mise en œuvre d'un système d'observation souvent complexe, le recueil d'opinions souffre d'une validité déficiente (Porter, 2011) et l'observation participante rend l'observateur juge et parti. Pour rendre l'observation de la pratique plus valide, il est bien sûr possible de la filmer et d'en réaliser une analyse *a posteriori*, aidée ou non par l'enseignant ayant mené la séance d'enseignement. Une assurance supplémentaire est d'avoir des instruments qui enregistrent avec une grande richesse de détails, dans leur multimodalité, les événements d'enseignement-apprentissage et les annotent automatiquement. De tels outils d'observation de l'enseignement en partie automatisés deviennent envisageables compte tenu des avancées dans le domaine du traitement et l'analyse du signal, et en apprentissage machine.

A titre d'exemple, il est couramment admis que tout enseignant adapte son enseignement au niveau de ses élèves, mais la manière dont cette adaptation est réalisée, et à partir de quels critères, a fait l'objet de nombreux travaux et débats (voir Wanlin & Laflotte, 2017 pour une synthèse). Lundgren (1972) a émis l'hypothèse que les élèves d'une classe peuvent être classés en sous-groupes de compétence dans chaque domaine, et que ces sous-groupes diffèrent dans leur compréhension des notions du domaine. Pour appréhender ce *delta*, l'enseignant s'appuie selon Lundgren sur un « groupe de pilotage » (*steering group*), et y adapte sa progression. L'étude des *steering groups* s'est réalisée jusqu'à présent à gros grain et de manière corrélationnelle. Il serait intéressant de les étudier en estimant ce signal : la fréquence des fixations du regard de l'enseignant sur ses élèves, *via* un oculomètre.

Depuis quelques années, des travaux se sont intéressés à capter et analyser l'expérience des enseignants et des élèves par oculométrie (voir Jarodzka *et al.*, 2020, pour une synthèse). Wolff *et al.* (2016) ont fait visionner à des enseignants du second degré novices et experts des vidéos de séances de cours comportant des événements problématiques, avec des élèves ayant des comportements plus ou moins rédhibitoires. L'analyse de la trajectoire du regard des enseignants sur la vidéo montre des *patterns* différenciant les novices et les experts (les experts scannent une plus grande part de l'image, sont plus centrés sur les endroits où l'information cruciale est disponible). Cortina *et al.* (2015) ont utilisé le paradigme expert-novice pour comparer la relation entre la qualité du climat de la classe et l'égalité de distribution des fixations du regard des enseignants sur les élèves de la classe. Ils ont montré que les enseignants experts distribuèrent leur regard plus équitablement que les novices et que cette équité était significativement corrélée avec la qualité des rétroactions fournies aux élèves par les enseignants.

Les moyens d'augmenter ces observations sont dorénavant nombreux : équiper enseignants et élèves de divers appareils mesurant leurs activités ouvertes (oculomètres, traceurs de positions, etc.) ou cachées (électroencéphalogramme, instrument de mesure de l'activité électrodermale, etc.). Ils se révèlent avantageux car, d'une part, une grande partie de l'activité d'enseignement-apprentissage est cachée (Nuthall, 2007) et, d'autre part, on ne peut analyser ces informations sur une durée raisonnable (*e.g.*, une heure) sans un travail intensif, ne serait-ce que pour rendre compte avec précision des déplacements d'un enseignant pendant un cours.

Les preuves en éducation et les moyens de les obtenir

Ces moyens d'observation répondent à un enjeu majeur : augmenter le nombre et la diversité de preuves à recueillir pour caractériser le plus finement et le plus écologiquement possible un événement scolaire, et *in fine* mieux analyser et documenter les pratiques des enseignants. Le courant de l'éducation fondée sur les données probantes part du principe que l'on peut tirer parti du nombre important de résultats de recherches scientifiquement prouvés pour que les acteurs du domaine (enseignants, personnels de direction, formateurs) prennent des décisions les mieux informées possible sur leur pratique (voir Saussez & Lessard, 2009 pour une synthèse critique). L'automatisation de leur recueil et de leur analyse grâce à des techniques d'intelligence artificielle est cependant ambivalente : d'un côté décupler le volume et la fiabilité des données probantes disponibles en condition écologique (*i.e.*, la plus proche des conditions réelles), d'un autre côté protéger cette écologie et ces protagonistes en déployant des garde-fous éthiques solides pendant la conception des systèmes automatisés, dans la mesure où leur puissance panoptique pourrait attenter à l'intégrité des individus.

Le format dans lequel présenter ces résultats peut être varié (*e.g.*, articles de recherche), mais un élément qui est rarement pris en considération est la nature des *exemples communs* (au sens de Kuhn, 1983) qui peuvent être présentés aux praticiens à des fins de réflexion dans leur résolution de problèmes pratiques. S'il est courant de présenter de tels exemples résolus dans l'enseignement des sciences, par exemple, cela ne l'est pas, ou très peu, dans le domaine de l'éducation. Notre propos est de « prélever » des exemples communs, recueillis de séances de classes *in vivo*, afin que les enseignants puissent les utiliser (que ce soit leurs séances ou celles de collègues) pour réfléchir à améliorer leur pratique (Bryk, 2017).

Un moyen simple, tiré de Claramunt *et al.* (1998), de caractériser ces exemples est de considérer *qui* (quel protagoniste), *où* et *quand* dans l'environnement considéré (la classe). Ces informations sont assez facilement récupérables par des moyens automatisés, et à partir d'elles, nous pouvons poser un cadre rendant compte de certaines actions de base que nous détaillerons dans les sections suivantes, ce cadre répond au *comment*. Un niveau

supplémentaire est accessible si nous disposons de théories explicatives en pédagogie, théories répondant au *pourquoi*.

L'apprentissage social selon Csibra et Gergely

Les données multimodales potentiellement récupérables d'un environnement de classe sont en nombre vraiment important : la classe est un environnement dynamique évoluant largement sans l'intervention de son superviseur (Rogalski, 2003) et les processus à l'étude sont en partie invisibles, donc nécessitant des inférences. Les preuves recueillies sont donc majoritairement indirectes.

Il est nécessaire de disposer d'un cadre théorique qui permette de les filtrer, de les mettre en correspondance avec des construits pédagogiques. De nombreux auteurs (voir Anderson & Burns, 1989 pour une synthèse) ont conçu des cadres descriptifs d'événements pédagogiques. Peu d'entre eux se sont intéressés aux événements d'enseignement basiques, où la communication parlée n'est pas présente ou joue un rôle mineur, et qui néanmoins peuvent entraîner un apprentissage. Le cadre que nous avons choisi est celui que Csibra et Gergely (2011) ont nommé « pédagogie naturelle », posant que la pédagogie est une activité sociale intégrée à la vie de tous les jours, et reposant sur certaines capacités humaines innées.

Ce cadre explique les habiletés d'enseignement et d'apprentissage social à partir d'habiletés de base comme l'attention conjointe, la communication et la réceptivité aux communications ostensives. Le principe général de ce cadre est de proposer des conditions nécessaires et suffisantes pour qu'il y ait enseignement, et ce cadre a été repris par de nombreux auteurs. Par exemple, Kline (2015) montre qu'une activité d'enseignement est nécessaire, d'une part lorsque l'apprenant rencontre deux types de difficultés : – avoir accès à un contenu, une connaissance, une information, une habileté ; – centrer son attention sur ce contenu ; et d'autre part lorsque l'enseignant peut allouer du temps pour aider l'apprenant, même si cela lui coûte.

L'analyse multimodale de situations d'enseignement

Nous utilisons ici la définition large de l'apprentissage social pour capturer et analyser des phénomènes se jouant entre l'enseignant et ses élèves en situation scolaire. Suivant Hoppitt et Laland (2008, p. 108), nous définissons l'apprentissage social comme « tout processus par lequel un individu (le démonstrateur)[l'enseignant] influence le comportement d'un autre individu (l'observateur)[l'élève] d'une manière qui augmente la probabilité que l'observateur apprenne ». Cela permet à des apprenants d'acquérir des informations plus efficacement que dans la condition où ils sont seuls, mais cela a aussi un coût pour les enseignants. Cette définition nous permet de nous centrer sur des événements tenus, parfois même non explicites et non intentionnels qui ne sont pas toujours considérés comme de l'enseignement en soi, mais qui l'étaient toutefois.

Une situation d'enseignement se perçoit de manière multimodale (réception d'informations *via* différents sens), dans un lieu fixe, avec une temporalité réduite (souvent, une ou deux heures). Cet article va ne présenter que des conditions d'enseignement ne nécessitant pas d'évaluation sémantique de la communication verbale (*i.e.*, n'impliquant pas la transcription et l'analyse du contenu de ce qui est communiqué). Les trois premiers niveaux présentés ci-dessous sont considérés comme de l'enseignement non-intentionnel (Gärdenfors & Högberg, 2021). Cette classification du processus le plus simple au plus complexe est nécessairement subjective, non hiérarchique et poreuse (Hoppitt & Laland, 2008). Nous postulons en outre que chaque niveau antérieur soutient et s'insère dans les niveaux ultérieurs.

Définitions

Définissons d'abord quelques termes qui vont permettre de rendre compte des différentes conditions d'apprentissage social.

- *Situation* : Un tel processus se déroule dans le temps et l'espace ; nous nommons situation une portion d'apprentissage social cohérente socialement et institutionnellement.
- *Condition* : Un type d'apprentissage social, impliquant de manière plus ou moins complexe et interreliée différentes habiletés de communication (principalement, mais pas uniquement) non verbales.
- *Champ* : Le champ est la partie de la situation d'enseignement/apprentissage directement accessible par les sens de l'enseignant et des élèves.
- *Information* : Le stimulus, au sens large, peut être un fait, une règle, une habileté, voire une technique que les apprenants peuvent imiter, ou encore en deviner des buts.
- *Présence à l'information et accès* : Kline (2015) a distingué le fait qu'un apprenant soit dans le même champ que l'information (simple présence), et le fait qu'il puisse y accéder de lui-même.

Présentation du cadre

Passons à la présentation des différentes conditions d'apprentissage social. Pour chaque condition, dérivée du cadre de Kline (2015), nous présentons ci-après le comportement de l'enseignant et des élèves, puis les outils principaux pour capturer et analyser des éléments de ce comportement, et enfin des études ayant mené à une implémentation en contexte scolaire de cette condition. La figure 1 décrit pour chacune de ces conditions : 0. la facilitation sociale, 1. la tolérance sociale, 2. la supervision sociale, et 3. l'amélioration sociale, les principaux flux interactionnels entre apprenants (A), enseignant (E) et le stimulus (S).

[Insérer Figure 1 ici]

0. Facilitation sociale ou la simple présence de l'enseignant

La première condition est celle de la simple présence de l'enseignant qui, *per se*, peut déjà influencer l'apprentissage, même si cela n'est pas systématique. L'enseignant est simplement présent, l'information, bien que présente et possiblement accessible, n'est pas ciblée par lui et il reste passif. Une étude surveillée est un exemple typique de cette condition. Dans cette condition, il n'y a pas d'enseignement intentionnel, mais cela induit un comportement spécifique des élèves. Être « en situation », à proximité des élèves, peut déjà être inducteur de certains comportements positifs : les élèves se savent à portée de l'observation ou de parole de l'enseignant et peuvent modifier leur comportement en conséquence, et donc apprendre des informations même si elles ne sont pas explicites. De plus, le fait qu'ils co-agissent augmente également la probabilité qu'ils apprennent ensemble (Zajonc, 1965). La proximité de l'enseignant avec les élèves a également été pointée comme une stratégie de gestion des comportements perturbateurs Gunter *et al.* (1995).

Les signaux multimodaux à récupérer de ce processus sont principalement des signaux de position spatiale et temporelle. Ces signaux sont perceptibles à l'aide de différents capteurs. Les technologies sans fil embarquées sur smartphones (*Bluetooth, Wifi, RFID*) peuvent servir à cette fin avec une précision de l'ordre du mètre dans des conditions optimales. Des capteurs de position spécialisés permettent d'obtenir une précision centimétrique. Si la scène est visible par plusieurs caméras calibrées sous plusieurs angles, il est possible de réaliser une reconstruction en trois dimensions des personnes et de leurs déplacements. Ces traitements automatiques produisent des annotations d'événements perçus dans les signaux multimodaux. Pour cette condition de facilitation sociale, ces annotations sont constituées de séquences de positions dans le temps.

Deux études récentes ont investigué cette direction. Lermigeaux-Sarrade (2018) a équipé des enseignants de sciences du second degré d'un dispositif permettant de les localiser précisément et a montré que la probabilité que les élèves s'engagent dans des activités hors-tâche augmentait dès lors que l'enseignant était à plus de 6 mètres des élèves concernés, et que la fréquence de sa visite dépassait les 30 minutes. Issaadi et Jaillet (2017) ont quant à eux réalisé des observations filmées en situation de classe. Ils ont montré que la proxémie pédagogique (le rapport de la pédagogie à des variables spatiales comme le positionnement, la distance ou la disposition) impactait significativement l'activité, les interactions sociales et le comportement d'étudiants, avec une inversion proportionnelle entre proximité de l'enseignant et échange entre étudiants, ainsi que l'existence d'espaces pédagogiques sociopètes vs sociofuges (favorisant la communication entre protagonistes vs la défavorisant), respectivement les tables de coin et de fond vs les tables proches de l'espace enseignant). Disposer des moyens qui renseignent cette proximité n'est toutefois qu'un premier pas, le second est de s'intéresser aux situations où l'information est présente dans le champ.

1. Tolérance sociale ou l'enseignant comme modèle

Les apprenants peuvent percevoir l'information à apprendre, mais son accès dépend de l'enseignant ; ce dernier « tolère » que les apprenants l'observent pendant des activités, où il donne involontairement l'exemple. Les apprenants regardent fréquemment l'enseignant, sans que le contraire soit vrai. Les postures et l'expression de l'enseignant sont ici importantes à capter (même si la communication peut déjà être verbale à ce niveau, nous ne nous y intéresserons pas). De nombreux exemples de cette condition peuvent être trouvés dans la classe : l'enseignant essuie le tableau, range son matériel, jette quelque chose dans la poubelle, ramasse et restitue un objet tombé à un élève, ou regarde incessamment son téléphone).

Deux concepts théoriques peuvent ici être mobilisés. D'une part, l'apprentissage vicariant (Bandura, 1980) ou apprentissage par observation d'un modèle, d'autre part le concept de proximité communicative (*immediacy*, voir Richmond *et al.*, 2003), c'est-à-dire le degré avec lequel un enseignant exprime des sourires, a un ton de voix expressif, est relaxé, fait que les élèves apprécient être là. Il augmente la possibilité par les apprenants de s'impliquer dans la situation et en venir à apprendre, sans qu'une transmission d'information en particulier puisse être réalisée intentionnellement par l'enseignant (Gorham, 1988). De cette activité peuvent découler des activités d'imitation, puisque les élèves seront d'autant plus à même de répéter les gestes, postures, expressions de l'enseignant dont ils auront été témoins.

Les signaux multimodaux à récupérer de ce processus d'apprentissage social sont ici plus nombreux. Les progrès récents en vision par ordinateur, utilisant des réseaux de neurones profonds (*deep learning*), permettent d'accroître l'éventail des informations perceptibles : variations d'activités enseignantes, expressions faciales et affects, postures, regards des apprenants vers l'enseignant, ainsi que nature des objets éventuellement médiateurs. La direction et la durée de fixation des regards peuvent désormais aisément être capturés par des oculomètres fixes ou portables. Plus parcimonieusement, les trajectoires des regards des apprenants peuvent être inférés et modélisés à l'aide de techniques d'intelligence artificielle en utilisant différents flux vidéo (Fang *et al.*, 2021) : les vues globales de la salle ou la vue à la première personne de l'oculomètre de l'enseignant. L'activité enseignante, de même que ses manifestations apparentes (expressions faciales, variation de posture, gestuelle) peuvent de même être extraites. La catégorisation des affects est plus difficile, mais dans un contexte de résolution de problèmes d'échecs il a été montré que les variations des affects sont pertinentes pour refléter certains états mentaux des participants Guntz *et al.* (2018). Enfin, les objets éventuellement médiateurs peuvent aussi être détectés (e.g., l'algorithme YOLO, Bochkovski *et al.*, 2020).

Les études automatisant la capture du comportement non verbal des protagonistes sont aujourd'hui très nombreuses, bien que peu aient été conduites pour investiguer cette condition spécifique. Les mesures et analyses automatisées d'expression faciale, de regard, de posture, de geste et d'activité, désormais robustes et calculables depuis un flux vidéo, sont encore peu exploitées. Bosch *et al.* (2016) ont exploré les possibilités de la vision par ordinateur pour détecter et prédire les affects des élèves à partir de leurs expressions faciales (canal primaire) et leurs mouvements corporels basiques (canal secondaire) pendant leurs interactions en classe. Bosch *et al.* (2018) ont utilisé les mêmes techniques en s'intéressant aux enseignants, et particulièrement à leurs mouvements et interactions avec le matériel pédagogique, afin d'évaluer leur soutien à l'attention des apprenants, et aussi leur activité non verbale, particulièrement en termes d'*immediacy*.

2. Procurer des opportunités aux apprenants et superviser le champ

Dans cette condition également, les apprenants pourraient être à même de percevoir l'information, mais leur accès à cette dernière est difficile et un aménagement d'une partie du champ est nécessaire. Ici, l'enseignant ne fait pas que tolérer passivement les apprenants, il aménage la situation pour que les apprenants puissent s'exercer, souvent par des activités de jeu, ou des tâches simples). L'enseignant ne supervise pas vraiment les apprenants, mais plutôt le champ dans son ensemble, qu'il a initié et aménagé (notamment *via* la planification et le suivi des activités). Il passe plus de temps à porter son attention sur le champ lui-même, afin de voir si la tâche est réalisée. Là encore, l'enseignant ne pointe pas spécifiquement sur l'information à acquérir. Les opportunités des deux précédents niveaux sont disponibles ici : la présence de l'enseignant et le fait qu'il tolère socialement les élèves renforcent la situation. Un exemple typique de cette condition ressort de la supervision d'une séance d'éducation physique et sportive, d'éducation musicale (*e.g.*, chorale) ou d'une séance de travaux pratiques (*e.g.*, pailleasse ou salle multimédia).

En complément des signaux des conditions précédentes, les signaux-clés pour appréhender cette condition sont la distribution du regard enseignant, l'activité et les regards des apprenants. Comme évoqué précédemment, ces informations sont en grande partie annotables automatiquement. Agrégées, elles peuvent dès lors permettre de répondre aux questions suivantes : quelle est l'activité jouée ? Quel est le comportement (non verbal) de l'enseignant pour superviser cette activité ? Quels sont les comportements apprenants subséquents ? Quelle est l'ambiance globale de la classe ?

Outre la reconnaissance automatique du type d'activité en jeu, le premier enjeu de telles analyses est d'analyser la supervision enseignante de la situation d'apprentissage proposée. Comme elle recourt surtout à sa vision, les études équipant l'enseignant d'un oculomètre mobile sont susceptibles de rendre compte de la distribution de son regard, de ses durées de fixation et de sa charge cognitive (voir Beach & McConnel, 2019 pour une revue). Un second enjeu pourrait être de mesurer la coopération des apprenants au champ planifié et supervisé par l'enseignant, étant entendu que la supervision enseignante et la coopération apprenante sont des processus aussi dynamiques que corrélés. Goldberg *et al.* (2019) ont utilisé des caméras puis l'apprentissage-machine pour analyser le comportement des élèves (*via* les suivis des regards, position de la tête et expressions faciales) et inférer l'engagement comportemental des apprenants. Ce type de recueil et d'analyse automatisée peut enfin contribuer à faire émerger des modèles de prédiction du climat de classe, coûteux en termes de cognition humaine (James *et al.*, 2019).

3. Enseigner par stimuli ou amélioration locale

Dans cette condition, les apprenants peuvent accéder à l'information mais manquent d'attention voire ne savent pas où regarder. L'enseignant expose l'élève à une information à

considérer, et/ou ralentit son action ou parole pour que cette information soit mieux perçue et comprise. Le concept de démonstration ostensive de Csibra et Gergely (2009) est ici particulièrement fécond. Ses manifestations les plus aiguës sont le contact visuel réciproque, même fugace, ainsi que la déictique. On peut s'attendre dans cette condition à des regards réciproques fréquents et à l'exagération de l'action de l'enseignant. Le contact oculaire a ici une double fonction : récupérer des informations pour superviser la situation, mais aussi communiquer. Csibra et Gergely (2009) expliquent notamment que cette démonstration ostensive est un vecteur de généralisabilité de l'objet appris (le référent) : les enfants s'attendent ainsi à apprendre quelque chose de généralisable dans des contextes ostensifs-référentiels plutôt qu'à simplement s'informer sur des faits épisodiques et particuliers qui n'existeraient que dans l'« ici et maintenant ». Cette gestuelle iconique peut également renvoyer à la pantomime de Gärdenfors (2017), structure de communication non verbale pendant lequel l'enseignant mime le mouvement d'une action relative à une tâche sans *réellement* l'accomplir, tout en s'assurant de l'attention et de la bonne perception de l'apprenant, le cas échéant en exagérant ou ralentissant certaines fractions d'actions.

De plus, des phénomènes d'attention partagée permettent de guider l'apprenant sur l'information, et sur les buts possibles de l'enseignant, ce qui amène un phénomène de synchronisation plus importante des activités de l'enseignant et des élèves. Cette condition est la première où nous traitons de l'enseignement en tant que tel, c'est-à-dire de processus intentionnels d'influence sociale de la part de l'enseignant (ce que Donker et ses collègues, 2020, nomment *agency*, par comparaison à *communion*). Csibra et Gergely (2009) avancent également que des signaux ostensifs existent dans la modalité auditive, par exemple, le modèle d'intonation particulier du « discours dirigé vers un enfant », appelé « *motherese* », ou « *mamanais* ». Cross (1989) transfère cette notion chez l'enseignant (« *teacherese* », ou « *ton professoral* ») qui souligne l'ostentation et adapte la prosodie enseignante aux signaux qu'il perçoit de la synchronie apprenante. Les exemples typiques de cette condition sont par exemple une séance de travaux dirigés, l'explicitation d'un graphique, un enseignant qui soutient une production d'élève, une démonstration conduite au tableau, etc.

Afin d'étudier cette condition, les signaux multimodaux déjà évoqués dans les conditions précédentes sont complétés par la prosodie de l'enseignant. Il s'agira ici d'analyser la temporalité des différents événements perçus du système (synchronie ou séquençement).

Un des enjeux importants de l'étude de cette condition est d'entrevoir la synchronie à l'œuvre au sein de la classe, entre enseignant et élèves, mais aussi entre les élèves. Cette synchronie serait notamment obtenue par des mécanismes d'attention partagée. Si Dikker *et al.* (2017) l'ont investiguée au moyen d'électroencéphalographes portatifs (*EEG headsets*), Fujii *et al.* (2018) proposent un *apparatus* moins contraignant, fondé là encore sur la vision par ordinateur et la reconnaissance automatique de l'inclinaison de la tête des participants. Fukuda *et al.* (2019) se sont intéressés à l'évaluation automatique de la prosodie afin d'analyser l'intention de l'enseignant au sein d'une salle de classe virtuelle, et il serait avantageux de croiser ces techniques de recueils à d'autres construits de la vie en classe (émotion, climat, interactions, synchronie, etc.).

Aspects éthiques

Capter, investiguer, voire rendre disponibles les analyses présentées dans les sections ci-dessus sur des événements se déroulant dans une classe n'est pas sans problèmes, même sous le sceau du RGPD (Règlement général sur la protection des données). Ces différentes captures de signaux sont en effet intrusives et gourmandes en données personnelles. La temporalité de la restitution de leur traitement automatisé tout autant que sa granularité interrogent également : doivent-elles alimenter un tableau de bord individuel à destination de l'enseignant

en temps réel (*online*), lui indiquant par exemple qui est inattentif, mécontent, dissipé, hors-tâche ? Cette configuration existe (voir par exemple les instrumentations décrites dans Lim *et al.*, 2017) mais nous semble singulièrement préjudiciable, d'une part à l'intégrité due aux participants à la classe, d'autre part à l'écologie de la classe. Si celle-ci devient omnisciente et omnipotente, nous postulons qu'elle induira en retour des comportements d'évitement, de dissimulation voire d'entrave, contre-productifs en termes de recueil de données probantes.

Deux garde-fous nous semblent en mesure de d'écarter ce péril, tels qu'exposés dans Laurent *et al.* (2020). Le premier utilise des techniques d'intelligence artificielle pour obfusquer (brouiller pour assurer l'anonymat) les données personnelles afin de ne saisir que leur « activité dans leur rôle » (enseignants, élèves, matériels) plutôt que leur biométrie ou sémantique. Le second est une précaution d'ingénierie de conception du dispositif automatisé dans laquelle l'éthique prévaut à tous les niveaux, notamment en prévoyant l'exclusion de toute rétroaction immédiate, pour lui préférer une restitution anonymisée, mais surtout différée et explicitée à tous les participants.

Discussion

Le cadre présenté a plusieurs avantages. Tout d'abord, il permet de caractériser automatiquement des types d'enseignement de complexité croissante et qui correspondent à la diversité de ce qu'on peut observer en situations authentiques. Ensuite, les moyens informatisés de traitement et d'analyse du signal utilisés sont relativement simples, puisque ne nécessitant pas d'analyse sémantique de la parole. Ces moyens produisent des annotations d'événements à différents niveaux, depuis leur perception directe jusqu'à leur analyse temporelle.

Si l'enjeu premier de ces modalités de recherche augmentée d'instrumentation automatisée et de techniques d'intelligence artificielle est bien de recueillir de nouvelles sources de données multimodales probantes dans des situations scolaires, il demeure conditionné à l'écologie du cadre dans lequel ces preuves sont recueillies. Aucun protagoniste ne doit être contraint ou trompé par les captures ou leur traitement, sous peine d'une part qu'on contrevienne à ses droits (comme professionnel ou élève), d'autre part de compromettre l'authenticité de la scène.

Ce cadre théorique permet aussi d'entreprendre de nouvelles études en psychologie et sciences de l'éducation en définissant des activités d'enseignement-apprentissage non ambiguës. Voici deux applications d'intérêt immédiat. Il est possible de capter et analyser certaines composantes de l'enseignement explicite Gauthier *et al.* (2007), où les phases d'explicitation de l'enseignant et d'entraînement des élèves pourraient être détectées, et aussi la distribution du regard de l'enseignant et ses interactions selon le genre des élèves (Jones & Dindia, 2004).

Remerciements

Ce travail a été partiellement financé par l'action Idex Formation 2018 (ANR 15-IDEX-0002 du PIA 2) de l'Univ. Grenoble Alpes.

Références

- Anderson, L. W., & Burns, R. B. (1989). *Research in classrooms, the study of teachers, teaching, and instruction*. Oxford: Pergamon Press.
- Bandura, A. (1980). *L'apprentissage social*. Bruxelles: Mardaga.

- Beach, P., & McConnel, J. (2019). Eye tracking methodology for studying teacher learning: a review of the research. *International Journal of Research & Method in Education*, 42(5), 485–501. doi: 10.1080/1743727X.2018.1496415
- Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint*. doi: arXiv:2004.10934
- Bosch, N., D'Mello, S. K., Baker, R. S., Ocumpaugh, J., Shute, V., Ventura, M., . . . Zhao, W. (2016). Detecting student emotions in computer-enabled classrooms. In G. Brewka (Ed.), *Proc. Twenty-Fifth Int. Joint Conf. on Artificial Intelligence (IJCAI-16)* (pp. 4125–4129). New York: AAAI.
- Bosch, N., Mills, C., Wammes, J. D., & Smilek, D. (2018). Quantifying Classroom Instructor Dynamics with Computer Vision. In C. P. Rosé, R. Martínez-Maldonado, H. U. Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren & B. d. Boulay (Eds.), *Proc. Int. Conf. AIED 2018* (pp. 30–42). Cham: Springer.
- Bryk, A. S. (2017). Accélérer la manière dont nous apprenons à améliorer. *Éducation et Didactique*, 11(2), 11–30. doi: 10.4000/educationdidactique.2796
- Claramunt, C., Parent, C., & Thériault, M. (1998). Design patterns for spatio-temporal processes. In S. Spaccapietra & F. Maryanski (Eds.), *Data mining and reverse engineering (DS-7)* (pp. 455–475). Dordrecht: Springer.
- Cortina, K. S., Miller, K. F., McKenzie, R., & Epstein, A. (2015). Where low and high inference data converge: Validation of CLASS assessment of mathematics instruction using mobile eye tracking with expert and novice teachers. *International Journal of Science and Mathematics Education*, 13(2), 389–403. doi: 10.1007/s10763-014-9610-5
- Cross, T. G. (1989). Teacher talk in preschool settings. *Early Child Development and Care*, 52(1-4), 133-146. doi: 10.1080/0300443890520111
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Science*, 13(4), 148–153. doi: 10.1016/j.tics.2009.01.005
- Csibra, G., & Gergely, G. (2011). Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society of London B*, 366, 1149–1157. doi: 10.1098/rstb.2010.0319
- Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., . . . Ding, M. (2017). Brain-to-brain synchrony tracks real-world dynamic group interactions in the classroom. *Current Biology*, 27(9), 1375–1380. doi: 10.1016/j.cub.2017.04.002
- Donker, M. H., van Gog, T., Goetz, T., Roos, A.-L., & Mainhard, T. (2020). Associations between teachers' interpersonal behavior, physiological arousal, and lesson-focused emotions. *Contemporary Educational Psychology*, 63. doi: 10.1016/j.cedpsych.2020.101906
- Fang, Y., Tang, J., Shen, W., Shen, W., Gu, X., Song, L., & Zhai, G. (2021). *Dual attention guided gaze target detection in the wild*. Paper presented at the IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- Fujii, K., Marian, P., Clark, D., Okamoto, Y., & Rekimoto, J. (2018). *Sync Class: Visualization System for In-Class Student Synchronization*. Paper presented at the Proc. 9th Augmented Human Int. Conf.

- Fukuda, M., Huang, H. H., & Nishida, T. (2019). *Detection of student teacher's intention using multimodal features in a virtual classroom*. Paper presented at the 11th International Conference on Agents and Artificial Intelligence (ICAART 2019), Prague.
- Gärdenfors, P. (2017). Demonstration and pantomime in the evolution of teaching. *Frontiers in Psychology*, 8. doi: 10.3389/fpsyg.2017.00415
- Gärdenfors, P., & Högberg, A. (2021). Evolution of intentional teaching. In N. Gontier, A. Lock & C. Sinha (Eds.), *The Oxford Handbook of Human Symbolic Evolution*. Oxford: Oxford University Press.
- Gauthier, C., Bissonnette, S., & Richard, M. (2007). L'enseignement explicite. In V. Dupriez & G. Chapelle (Eds.), *Enseigner*. Paris: PUF.
- Goldberg, P., Sümer, Ö., Stürmer, K., Wagner, W., Göllner, R., Gerjets, P., . . . Trautwein, U. (2019). Attentive or Not? Toward a Machine Learning Approach to Assessing Students' Visible Engagement in Classroom Instruction. *Educational Psychology Review*, 33(1), 27–49. doi: 10.1007/s10648-019-09514-z
- Gorham, J. (1988). The relationship between verbal teacher immediacy behaviors and student learning. *Communication Education*, 37(1), 40-53. doi: 10.1080/03634528809378702
- Gunter, P. L., Shores, R. E., Jack, S. L., Rasmussen, S. K., & Flowers, J. (1995). Using teacher/student proximity to improve students' behavior. *Teaching Exceptional Children*, 28(1), 12–14. doi: 10.1177/004005999502800103
- Guntz, T., Balzarini, R., Vaufreydaz, D., & Crowley, J. (2018). Multimodal observation and classification of people engaged in problem solving: Application to chess players. *Multimodal Technologies and Interaction*, 2(2), 11. doi: 10.3390/mti2020011
- Hoppitt, W., & Laland, K. N. (2008). Social Processes Influencing Learning in Animals: A Review of the Evidence. *Advances in the Study of Behavior*, 38, 105–165. doi: 10.1016/s0065-3454(08)00003-x
- Issaadi, S., & Jaillet, A., (43). (2017). Proxémie d'apprentissage. Éducation et socialisation. *Les Cahiers du CERFEE*, 43. doi: 10.4000/edso.1960
- James, A., Chua, V. Y., Maszczyk, T., Nunez, A. M., Bull, R., Lee, K., & Dauwels, J. (2019). Automated Classification of Classroom Climate by Audio Analysis. In L. D'Haro, R. Banchs & H. Li (Eds.), *9th International Workshop on Spoken Dialogue System Technology* (pp. 41–49). Singapore: Springer.
- Jarodzka, H., Skuballa, I., & Gruber, H. (2020). Eye-Tracking in Educational Practice: Investigating Visual Perception Underlying Teaching and Learning in the Classroom. *Educational Psychology Review*. doi: 10.1007/s10648-020-09565-7
- Jones, S. M., & Dindia, K. (2004). A Meta-Analytic Perspective on Sex Equity in the Classroom. *Review of Educational Research*, 74(4), 443–471. doi: 10.3102/00346543074004443
- Kline, M. A. (2015). How to learn about teaching: An evolutionary framework for the study of teaching behavior in humans and other animals. *Behavioral and Brain Sciences*, 38, 1-71. doi: 10.1017/s0140525x14000090
- Kuhn, T. S. (1983). *La structure des révolutions scientifiques*. Paris: Flammarion.
- Laurent, R., Vaufreydaz, D., & Dessus, P. (2020). Ethical teaching analytics in a Context-Aware Classroom: A manifesto. *ERCIM News*, 120, 39–40.

- Lermigaux-Sarrade, I. (2018). *Rôle de l'organisation de l'espace de travail dans les activités effectives et empêchées des enseignants*. (Thèse de sciences de l'éducation), Université Grenoble Alpes, Grenoble.
- Lim, J. H., Teh, E. Y., Geh, M. H., & Lim, C. H. (2017). Automated classroom monitoring with connected visioning system. In W.-S. Gan, K. Wong & K. Ma (Eds.), *2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC 2017)* (pp. 386–393). Hong Kong: IEEE.
- Lundgren, U. P. (1972). *Frame factors and the teaching process*. Stockholm: Almqvist & Wiksell.
- Nuthall, G. (2007). *The hidden lives of learners*. Wellington: NZER Press.
- Porter, S. R. (2011). Do college student surveys have any validity? *The Review of Higher Education*, 35(1), 45–76. doi: 10.1353/rhe.2011.0034
- Richmond, V. P., McCroskey, J. C., & Johnson, A. D. (2003). Development of the Nonverbal Immediacy Scale (NIS): Measures of the self- and other-perceived nonverbal immediacy. *Communication Quarterly*, 51(4), 504–517.
- Rogalski, J. (2003). Y a-t-il un pilote dans la classe ? Une analyse de l'activité de l'enseignant comme gestion d'un environnement dynamique ouvert. *Recherche en Didactique des Mathématiques*, 23(3), 343–388.
- Saussez, F., & Lessard, C. (2009). Entre orthodoxie et pluralisme, les enjeux de l'éducation basée sur la preuve. *Revue française de pédagogie*, 168, 111-136. doi: 10.4000/rfp.1804
- Wanlin, P., & Laflotte, L. (2017). Connaissances des enseignants sur le(ur)s élèves et leur utilisation pour donner cours. Genève: Université de Genève, rapport de recherche 150316.
- Wolff, C. E., Jarodzka, H., van den Bogert, N., & Boshuizen, H. P. A. (2016). Teacher vision: Expert and novice teachers' perception of problematic classroom management scenes. *Instructional Science*. doi: 10.1007/s11251-016-9367-z
- Zajonc, R. B. (1965). Social facilitation. *Science*, 149(3681), 269–274.

Figure 1 : les quatre conditions d'apprentissage social, dérivées de Kline (2015).

