



Laboratoire d'Informatique de l'Université du Maine

Projet PASTEL

Performing Automated Speech Transcription for Enhancing Learning

D4.2

Transcription Automatique et Traces dans un EIAH – État de l'Art

Date de dépôt prévisionnelle Novembre 2017

Révision 1 du 16 Mars 2018

Auteur principal : Vincent Barré – LIUM

Contributeurs : Madeth May – LIUM



Laboratoire d'Informatique de l'Université du Maine

Table des matières

1.	Transcription Automatique et traces dans un EIAH.....	3
1.1.	Utilisation de la Transcription Automatique en contexte d'apprentissage.....	5
1.1.1.	Enrichissement du contenu pédagogique.....	6
1.1.2.	Assistance / Remédiation.....	9
1.1.3.	Évaluation des apports de la TA.....	9
1.1.4.	Évaluation des apprenants.....	10
1.1.5.	Outillage de la collaboration.....	10
1.2.	Apports de la TA temps-réel en contexte d'apprentissage.....	11
1.3.	Traces et indicateurs.....	12
2.	Bibliographie.....	15

1. Transcription Automatique et traces dans un EIAH

Cet état de l'art s'inscrit dans le cadre du projet PASTEL (*Performing Automated Speech Transcription for Enhancing Learning*) qui a pour objectif d'explorer le potentiel de la transcription automatique (TA) en temps réel pour l'instrumentation de situations pédagogiques mixtes, où les modalités d'interaction sont présentes ou à distance, synchrones ou asynchrones.

Dans le cadre du projet PASTEL, nous nous proposons d'étudier et d'instrumenter deux situations pédagogiques : un exposé oral de type cours magistral synchrone, présentiel ou à distance et un travail collaboratif de type travaux pratiques synchrone, présentiel ou à distance.

Cet état de l'art sera donc structuré autour de ces deux situations typiques résumées ci-après et recensera les usages de la TA et leur contexte.

Situation 1 : cours magistral synchrone, présentiel ou à distance			
Source TA : transcription du discours de l'enseignant			
<i>Usage</i>		<i>Limites</i>	<i>Effets positifs</i>
Sous-titrer l'intervention	Transcription semi-automatique	Bonne qualité mais coût potentiellement élevé	Effet positif malgré les inconvénients pour certains publics en situation de handicap ou non natifs, travaux sur l'automatisation de la correction de la transcription
	Transcription automatique	Qualité non optimale, phase d'entraînement	
Traduire l'intervention			
Indexation des captations du discours de l'enseignant	Chapitrage	Contenu textuel des diapos meilleur indicateur de sujet que TA	
	Typage des séquences		

	Enrichissement via des banques de ressources externes		Peut permettre une personnalisation des apprentissages
Aide pour prise de notes		Bonne qualité nécessaire, notamment sur vocabulaire spécifique, besoin de formation pour concevoir et utiliser à bon escient des notes	TA utile pour améliorer la compréhension du cours et enrichir les interactions avec l'enseignants ou les pairs
Environnement combinant discours, diaporama et tableau			Effet positif sur apprenants ayant des besoins spécifiques ou n'ayant pu assister au cours
Assistance / remédiation	Accessibilité, difficultés physiques ou cognitives		Effets étudiés et positifs sur publics spécifiques

Situation 2 : travaux pratiques synchrone, présentiel ou à distance			
TA : transcription des échanges			
<i>Usages</i>		<i>Limites</i>	<i>Effets positifs</i>
Contexte multi-locuteurs		Baisse de la qualité de TA, en particulier si unique micro d'ambiance	Possible d'équiper chaque locuteur ou d'utiliser un intermédiaire qui reformule
Contexte multi-sites		Problèmes de débits et/ou d'interférences lors des prises de parole	TA améliore la qualité globale des échanges et la performance du groupe

Contexte multiculturel			TA et traduction automatique permettent un travail en commun
Contextualisation de la TA	Reconnaissance locuteur		
	Détection de rôles		
	Détection des changements de sujets de conversation		
Évaluation des apprenants	Analyse de la prononciation (langues étrangères ou maternelle pour personnes malentendantes)		Permet une auto-évaluation et augmente l'autonomie des apprenants

1.1. Utilisation de la Transcription Automatique en contexte d'apprentissage

Si l'évolution des technologies de transcription automatique rendent désormais envisageable un usage temps réel en contexte d'apprentissage, les usages en temps plus ou moins différé se sont d'ores et déjà développés depuis une quinzaine d'années.

Au tournant des années 2000, la TA fut d'abord utilisée pour accompagner des publics spécifiques (apprenants ayant des déficiences physiques ou cognitives, apprenants non natifs...), puis fut ensuite utilisée plus largement dans des environnements d'apprentissage traditionnels ou en ligne pour enrichir des activités par la transcription de cours magistraux ou pour le soutien aux activités collaboratives [1].

Cet usage croissant de la transcription automatique dans divers environnements d'apprentissage a conduit à une réflexion plus générale sur l'inclusion des technologies de TA dans les processus pédagogiques. Le projet Européen Net4Voice¹ s'est par exemple attaché à définir comment la TA pouvait être un moyen d'aider les enseignants et les apprenants dans leurs processus d'enseignement/apprentissage (que ce soit pour des apprenants du secondaire, du supérieur ou en FTLV) [2].

Cette première section abordera l'usage de la TA pour différentes natures d'activités (enrichissement du contenu pédagogique, assistance / remédiation, évaluation et outillage de

¹ <http://www.net4Voice.eu>

la collaboration) pour lesquelles différents contextes et/ou publics seront étudiés (apprenants ayant des difficultés particulières, apprentissage des langues...). La question du temps réel / différé sera, quant à elle, abordée plus en détail dans la prochaine section avant de terminer cette première partie de l'état de l'art par la problématique des traces et de la production d'indicateurs dans un contexte d'usage de la TA.

1.1.1. Enrichissement du contenu pédagogique

Un scénario habituel d'usage de la TA en contexte pédagogique est souvent celui du cours magistral « classique » pour lequel l'intervention orale de l'enseignant (accompagnée du diaporama utilisé et éventuellement d'autres supports de cours) constitue l'essentiel du contenu pédagogique. C'est sur cette intervention orale de l'enseignant que la TA sera appliquée pour enrichir le contenu pédagogique [3].

Un premier type d'enrichissement, naturel même si complexe d'un point de vue technologique, consiste à sous-titrer, transcrire, voire traduire en d'autres langues la parole de l'enseignant. Si les usages asynchrones peuvent bénéficier d'un traitement semi-automatique et donc d'une qualité de transcription tout à fait acceptable [4] [5], ce n'est pas le cas des usages synchrones pour lesquels la qualité de la transcription automatique (sans intervention humaine) reste un frein [6]. Cette qualité de transcription peut néanmoins être améliorée via des phases d'entraînement du système et du locuteur avant de débiter le cours [7], mais cet entraînement préalable constitue également un frein important à la diffusion de la TA en contexte pédagogique.

Ces différents freins peuvent néanmoins plus facilement être surmontés lorsque le bénéfice de la TA pour les apprenants est clair. C'est par exemple le cas pour répondre aux besoins des apprenants non natifs ou ayant des besoins particuliers (apprenants malentendants par exemple) pour lesquels un effet positif est observé [8].

Au-delà des publics particuliers, l'intérêt de la TA pour améliorer les apprentissages dans un environnement de classe synchrone en ligne a été questionné par de nombreux auteurs. L'introduction de la TA passe par une phase d'appropriation de la technologie, tant pour l'enseignant qui doit faire attention à préserver un bon taux de reconnaissance, que pour les apprenants qui doivent apprendre à utiliser les transcriptions. Après cette phase d'appropriation, un effet positif de la TA est observé [9] lorsque les apprenants utilisent le texte généré pour améliorer leur compréhension du cours, à la fois durant (temps réel) et après (temps différé) le cours.

Cet usage renvoie également aux besoins identifiés dans le lot 1 du projet PASTEL [10], et en particulier que la production de sous-titres en temps réel (via la TA) ne ressort pas comme un élément central (pour un public « classique »), mais néanmoins utile après coup pour diverses utilisations (livrable du lot 1, p. 14).

Pendant les sessions d'apprentissage, les apprenants préfèrent en effet généralement écouter l'orateur tout en lisant la transcription lorsqu'elle est disponible en temps réel, permettant ainsi de clarifier et d'affiner leur compréhension du discours de l'enseignant [11]. Les

apprenants indiquent également que la transcription aide pour la prise de note en séance ou lorsqu'ils retravaillent le cours après la séance.

Une fois transcrit (idéalement avec une qualité suffisante), le verbatim de la parole de l'enseignant peut ensuite être utilisé de nombreuses façons pour enrichir le matériel pédagogique résultant de la captation du cours. Les vidéos de cours peuvent ainsi être indexées automatiquement en fonction du discours de l'enseignant et même segmentées automatiquement en fonction du déroulé des séquences du cours.

Les vidéos de cours sont en effet une source importante d'informations pour les apprenants, à condition que ces vidéos soient chapitrées en éléments cohérents et indexées pour permettre des recherches thématiques. Un premier moyen de chapitrer les vidéos est d'utiliser le tempo imprimé par le diaporama de l'enseignant et de découper la transcription en séquences correspondants aux diapositives [12]. Le contenu textuel de la diapositive permet également d'indexer le contenu du segment, éventuellement renforcé par la transcription du discours de l'enseignant [13] (mais le contenu textuel de la diapositive est généralement un meilleur indicateur que la transcription pour détecter le sujet de la séquence [14]).

Au-delà du chapitrage en lien avec les diapositives, il est également possible de typer les séquences pour faire apparaître la structure du discours (ou métadiscours) sous forme d'introduction, de conclusion, etc. [15].

La transcription de la parole de l'enseignant est par ailleurs un matériau de choix pour accompagner les apprenants dans leur prise de note ou leur fournir en fin de séance un résumé automatique des concepts abordés.

L'aide à la prise de note (collaborative ou non) en lien avec le cours est l'un des premiers usages explorés de la TA en contexte de formation [16]. Même si les apprenants ont des pratiques variées en la matière, et ont généralement besoin d'une formation pour concevoir et utiliser à bon escient ces notes [17], l'usage de la TA pour concevoir des notes requiert une bonne qualité de transcription, en particulier sur le vocabulaire spécifique [18].

L'amélioration de l'aide à la prise de note et à l'usage des textes issus de la transcription automatique passe en particulier par l'analyse des stratégies mises en œuvre par les apprenants avec ces textes « aide-mémoire ». Dans [19], 3 grandes stratégies d'usage de la transcription ont été identifiées parmi des apprenants ayant reçu les textes transcrits d'un cours auquel ils avaient participé : (a) comprendre le contenu, (b) trouver des réponses et (c) écrire un résumé.

Les transcriptions sont généralement utiles pour améliorer la compréhension du contenu de l'enseignement [20] et permettent aux apprenants de s'y référer lorsqu'ils répondent ou posent des questions à l'enseignant ou à leurs pairs.

L'identification des concepts au sein de la transcription d'un enseignement peut également permettre d'enrichir la séquence en proposant automatiquement (en temps réel ou différé) des ressources documentaires annexes ou encore l'approfondissement des concepts abordés (définition, ressources complémentaires, etc.).

De ce point de vue, l'indexation de vidéos à l'aide de la transcription automatique de la bande son et de la reconnaissance des textes écrits qu'elles contiennent offre l'opportunité de fouiller de grandes banques de vidéos à la recherche de notions / concepts spécifiques [21] en vue d'enrichir la séquence pédagogique.

Un enseignement est généralement conçu pour un « apprenant-type », alors que chaque apprenant a plutôt besoin de ressources appropriées à son style d'apprentissage et à ses propres besoins de remédiation. Cette problématique est particulièrement importante dans le cadre des MOOCs au regard de l'hétérogénéité du public, mais se retrouve plus largement dans tout type d'enseignement. Cette question a, par exemple, été abordée par [22] qui proposent un cadre de référence (*educational content linking*) permettant de lier et d'organiser les ressources fournies par le concepteur, mais aussi pour proposer des ressources alternatives aux apprenants. L'ordonnancement des ressources et l'identification des liens de précédences entre elles pouvant alors être automatisés via des approches de type *machine learning* [23].

Cette problématique de l'enrichissement avec des ressources externes a été abordée lors du quick-off du projet PASTEL et figure dans l'analyse des besoins réalisée dans le lot 1 [10] :

- *Accès direct à des contenus pédagogiques présentant les concepts principaux présentés lors des CM en dehors du cours et pendant les TD*
- *Mise à disposition, organisation et hiérarchisation de contenus (ressources en ligne, diaporama, etc.) venant enrichir le contenu du CM*
- *Création d'un glossaire relatif au contenu du cours*
- *Possibilité d'annoter le contenu du cours et de partager ces annotations, mise en évidence pour l'enseignant des annotations correspondant aux demandes d'explication ou d'approfondissement*
- *Transcription de vidéos pédagogiques connexes permettant la mise à disposition de ressources complémentaires sur les thèmes traités*

Dans le cadre du projet PASTEL, la synchronisation du diaporama, du tableau et du discours de l'enseignant sera explorée.

Le potentiel de la transcription automatique du cours étant globalement avéré (dès lors que la qualité de la transcription est suffisante), de même que l'intérêt du diaporama pour la structurer et l'indexer, il est alors naturel de vouloir chercher à combiner diaporama, discours et informations écrites au tableau pour fournir un tout cohérent aux apprenants [24]. Les apprenants ayant des besoins spécifiques, ceux n'ayant pu assister au cours ou encore ceux ayant des difficultés à prendre des notes tout en écoutant, regardant et réfléchissant constituant les cibles premières [25].

Plusieurs systèmes implémentent d'ores et déjà une telle synchronisation diaporama / tableau / discours, mais en temps différé (par exemple [26]). Une autre voie explorée dans le projet COCo² fut l'alignement multimodal diaporama / vidéo / notes collaboratives, même si les notes n'étaient pas basées sur la transcription automatique de la vidéo [27].

² <http://www.comin-ocw.org/>

1.1.2. Assistance / Remédiation

Les premiers usages de la TA en contexte éducatif ont d'abord été tournés vers les apprenants ayant des besoins spécifiques [28], que ce soit des apprenants malentendants ou ayant des difficultés de compréhension (apprenants non natifs par exemple), afin de leur permettre de suivre « normalement » les enseignements et de répondre ainsi aux critères institutionnels d'accessibilité [29] [3].

Au-delà des usages tournés vers les apprenants malentendants ou non natifs, d'autres types de besoins spécifiques peuvent bénéficier des apports de la TA pour faciliter l'intégration. C'est par exemple le cas des apprenants ayant des difficultés avec l'orthographe et l'écriture (troubles dys), pour lesquels la modalité de saisie orale est plus adaptée que le clavier [30], mais aussi des déficients visuels [31] pour lesquels des systèmes plus spécifiques de TA, comme les systèmes « *Voice-to-Braille* », sont utilisés.

Plus récemment, de nouveaux systèmes mêlant TA et réalité augmentée permettent d'étendre les possibilités de la TA à destination de publics spécifiques, par exemple pour les malentendants [32].

Pour tous ces usages, la question de la qualité de la captation sonore en environnement de classe se pose, d'autant plus qu'elle influe de façon importante sur la qualité de la transcription et la facilité d'utilisation en environnement réel [33], où de multiples locuteurs peuvent s'exprimer.

Pour des publics ayant des besoins particuliers, la TA peut être un élément central de la mise en place d'environnements d'apprentissage, comme pour les enfants ayant un trouble du spectre de l'autisme, pour lesquels la TA peut contribuer à constituer un environnement d'apprentissage complet [34], ou encore pour l'apprentissage du braille pour les déficients visuels [35].

1.1.3. Évaluation des apports de la TA

Les systèmes de TA étant utilisés dans un contexte d'apprentissage, il est naturel de chercher à en analyser l'effet sur les apprentissages. Dès les origines de la TA, la qualité (taux de reconnaissance du discours) de la transcription est apparue comme étant un facteur clé. Son influence concrète s'est néanmoins révélée plus ou moins importante selon le public concerné et ses attentes vis-à-vis de la TA.

Au-delà de la conception de dispositifs de TA et de situations pédagogiques mettant en œuvre ces dispositifs, la question de l'évaluation de l'usage de ces dispositifs est cruciale et permet d'en apprécier, ou non, l'efficacité. Cette évaluation doit porter sur l'ensemble des situations, collaboratives [36] ou non, pour des publics spécifiques ou non, dans lesquelles la TA est utilisée.

Le public des apprenants non natifs est l'un de ceux qui ont été le plus étudié, en particulier lors de cours en Anglais pour lesquels un effet bénéfique sur les non natifs est observé [37] [38] [39] [40].

Des effets faiblement positifs ont également été observés pour des adolescents et adultes ayant des difficultés d'apprentissage [41].

Au-delà de publics particuliers, plusieurs projets nationaux ont cherché à évaluer le potentiel de la TA en contexte de classe, que ce soit dans l'enseignement primaire [42], secondaire ou supérieur [43] et pour un public diversifié.

1.1.4. Évaluation des apprenants

Les technologies de transcription automatique de la parole peuvent participer à l'(auto)évaluation des apprenants, en particulier dans le domaine des langues où ces technologies sont généralement utilisées pour analyser la prononciation des langues étrangères et augmenter l'autonomie des apprenants [44] [45] [46] [47].

Elles sont également utilisées pour l'apprentissage de la parole chez les personnes malentendantes [48].

Au-delà de ces cas particuliers, peu d'usages de la transcription automatique sont recensés pour contribuer à l'(auto)évaluation des apprenants.

1.1.5. Outillage de la collaboration

L'usage de systèmes de TA dans un contexte multi-locuteurs induit généralement une baisse de la qualité de la transcription (en particulier lors d'une captation avec un unique micro d'ambiance), même s'il est par exemple possible, pour pallier à cette difficulté, de revenir à une situation avec un unique locuteur en faisant reformuler les questions par un animateur [3] ou d'équiper chaque locuteur d'un micro (individualisant ainsi les flux audio tout en améliorant la qualité de la captation).

Dans une situation où plusieurs intervenants s'expriment, la question de la reconnaissance du locuteur [49] et de la détection des rôles tenus par les intervenants se pose pour contextualiser la transcription.

En vue d'outiller la collaboration au sein d'un groupe d'apprenants, il peut, par exemple, être intéressant de savoir détecter le rôle (qu'il soit social ou orienté vers une tâche) tenu par chaque participant et éventuellement le faire évoluer. La reconnaissance de ce rôle [50] peut, par exemple, se baser sur le champ lexical utilisé par chaque participant, sur le réseau des interactions entre participants [51] ou encore sur une combinaison de ces différents indicateurs [52].

Un cas particulier de la détection de rôle consiste à détecter l'émergence de leaders dans un groupe pour les mettre en évidence [53].

Lors de conversations de groupe, au-delà des locuteurs et des rôles qu'ils tiennent, il est également intéressant de pouvoir détecter les changements de sujet de conversation [54], par exemple pour pouvoir ensuite analyser le déroulement de la séance.

Dans le cadre de travaux collaboratifs de groupe (interculturalité, synchrone, online / traditionnel, etc.) la TA peut servir à faciliter la communication entre les membres et aider à synthétiser les échanges.

Dans un contexte de classe en ligne où les communications inter-sites sont parfois dégradées à cause du débit et/ou des interférences lors de la prise de parole, la transcription automatique des échanges peut améliorer la qualité globale des échanges et la performance du groupe [55] [56].

Dans un cadre interculturel, la transcription automatique associée à la traduction automatique permet de faire travailler ensemble des apprenants ne partageant aucune langue en commun avec un taux de satisfaction suffisant [57].

Plusieurs pistes d'usage de la TA dans un tel contexte ont également émergé lors du kick-off du projet PASTEL et en particulier la régulation de la prise de parole (aider les apprenants à s'autoréguler lors de la prise de parole dans un groupe en mettant en évidence les leaders et ceux qui interviennent le moins), la mise en place d'outils de tutorat pour réguler le suivi par et pour les apprenants (vérifier les dérives, hors sujet...) ou le développement de liens intergroupes (établir des liens basés sur l'activité et permettre des comparaisons entre les groupes).

Plusieurs besoins ont également été mis en évidence dans le cadre du lot 1 du projet PASTEL [10], en particulier le besoin d'outils favorisant le partage, la communication et l'échange autour des contenus pédagogiques, la possibilité de solliciter plus directement l'enseignant pendant les cours magistraux pour exprimer un besoin de clarification de certains concepts ou encore la présentation d'indicateurs synthétiques (progression, difficultés...) portant sur le travail des apprenants lors des travaux de groupe.

1.2. Apports de la TA temps-réel en contexte d'apprentissage

La transcription en temps réel de l'intervention d'un enseignant lors d'un cours peut permettre d'en favoriser l'accessibilité à tous les publics. Dans des cas particuliers, elle s'effectue manuellement, à l'aide d'opérateurs sténographes qui réalisent cette transcription en temps réel. Cela reste néanmoins réservé à des situations très spécifiques et n'est pas généralisable pour des raisons de coût. La question de la transcription automatique en temps réel s'est donc très rapidement posée, dès l'introduction d'outils capables de la réaliser.

La transcription automatique temps réel (et même en temps différé) a longtemps posé de sérieux problèmes liés à la qualité de la transcription (souvent insuffisante pour être réellement utilisable dans un contexte pédagogique), à l'entraînement nécessaire du système et du locuteur, à la qualité de la prise de son, au besoin de disposer du vocabulaire du domaine pour entraîner le système, etc. Les comparaisons entre transcription temps réel et transcription différée (enrichissement par sous-titrage d'une vidéo mise à disposition) en conditions d'apprentissage montrent d'ailleurs clairement l'importance de la qualité du sous-titrage [58].

De nombreux travaux se sont alors concentrés sur l'automatisation de la correction de la transcription produite par les systèmes (temps réel ou non) pour éditer et corriger les erreurs de sous-titrage [59] [60] [61].

Plus récemment, la puissance de calcul et les algorithmes de TA permettent désormais une transcription de relativement bonne qualité, y compris avec des moyens « individuels » (sans nécessiter de puissance de calcul distante) [62].

1.3. Traces et indicateurs

La plupart des travaux concernant l'efficacité pédagogique de la transcription automatique reposent sur des méthodes qualitatives basées sur des interviews et des questionnaires. Peu de travaux explorent des méthodes quantitatives, à l'instar de [40] qui utilisent des méthodes d'*eye-tracking* pour analyser l'usage des textes transcrits par les apprenants. Néanmoins, les techniques de *Learning Analytics* (LA) peuvent apporter d'intéressants éclairages.

En effet, les traces d'interactions peuvent être exploitées dans les situations d'apprentissage utilisant la TA pour aider à la fois les chercheurs à mieux concevoir le système et les enseignants à identifier les difficultés des apprenants (qu'elles soient relatives à la compréhension des concepts abordés dans le cours, au besoin de revenir sur les concepts déjà abordés, ou encore au rythme du cours trop rapide, etc.).

Les techniques de *Learning Analytics* [63] permettant la collecte, l'analyse et la visualisation des indicateurs calculés à partir des traces, sont ainsi associées au contexte pédagogique instrumenté par la TA. Elles concernent l'identification des besoins d'observation, la définition des indicateurs d'interactions, la construction des indicateurs et de leur visualisation. Les travaux de [64] se focalisent sur l'observation en temps réel des interactions des apprenants sur une plateforme de formation en ligne (*e.g.* Moodle) et sur la visualisation des indicateurs. Ceux de [65] montrent également l'importance de la visualisation d'indicateurs pour analyser les interactions entre les apprenants et le système et fournir une information enrichie aux tuteurs. Ces travaux, comme de nombreux autres, démontrent l'utilité d'exploiter les traces d'interactions des apprenants afin de fournir un véritable service aux enseignants-tuteurs souhaitant suivre en temps réel ou différé les activités des apprenants, mais aussi l'utilité de les exploiter à des fins de réingénierie à l'issue de la session de formation.

L'état de l'art sur les technologies d'assistance dans un cadre d'apprentissage collaboratif proposé par [66] met en avant l'importance de collecter les traces d'interaction et d'en faciliter l'interprétation, tant par les participants, apprenants ou tuteurs, que par le système lui-même. Les auteurs proposent également une classification de ces systèmes selon le degré de maturité des indicateurs mis à disposition des acteurs. Bien évidemment, l'intérêt de la collecte et de l'analyse des traces d'interactions n'est pas limité au domaine du CSCL comme le montrent, par exemple, [67] dans une étude portant sur l'utilisation de systèmes de traçage dans diverses situations. Ces travaux concluent, en particulier, que l'assistance apportée par la mise à disposition de ces indicateurs est bénéfique sur les apprentissages et l'assimilation des contenus.

L'état de l'art proposé par [68] dans le cadre du NoE Kaléidoscope³ (et en particulier dans le cadre du projet ICALTS⁴) pose plus généralement le cadre de l'analyse des interactions en vue de soutenir l'autorégulation des tuteurs et des apprenants et porte sur la définition d'indicateurs et d'outils liés à l'analyse des interactions et de la collaboration.

L'analyse des interactions vise à fournir des informations pertinentes directement aux apprenants et/ou aux tuteurs dans le but d'autoévaluer leur activité. Cette analyse est en particulier fondamentale pour analyser et comprendre les interactions très complexes qui se tiennent dans les systèmes collaboratifs tels que celui mis en œuvre dans la seconde situation qui sera abordée dans le projet PASTEL.

Les techniques de *Social Learning Analytics*, basées sur l'analyse des interactions sociales dans un processus d'apprentissage, sont particulièrement intéressantes pour les travaux de groupe ou les échanges au sein de forums [69].

Les « indicateurs » résultant de l'analyse des interactions doivent être mis à disposition des acteurs de la situation d'apprentissage sous un format approprié (graphique, numérique, littéral...) pouvant être interprété par les apprenants ou les tuteurs qui peuvent ainsi avoir un accès réflexif sur leurs activités. Ces « indicateurs » peuvent éclairer sur la « qualité » des activités individuelles, le type ou la qualité de la collaboration, la qualité des produits de la collaboration ou encore l'appropriation de leur processus de production.

Par exemple, pour [70], les interactions au sein d'un environnement de type CSCL sont analysées en vue de quantifier la participation des apprenants et donner une indication sur leurs chances de succès dans l'activité via des algorithmes génétiques.

Plus généralement, de nombreux travaux mettent en avant les services innovants que ces systèmes de traçage apportent à l'ensemble d'acteurs d'apprentissage, y compris les concepteurs-développeurs, les chercheurs, les enseignants, les apprenants, etc.

Les indicateurs jouent ainsi un rôle crucial dans le processus d'apprentissage, à savoir le suivi en temps réel d'une activité pédagogique, l'autorégulation, l'analyse des comportements d'un individu ou d'un groupe d'apprenants...

La conception et l'implémentation de mécanismes de collecte des traces et de calcul d'indicateurs est donc importante en vue d'aider les apprenants à prendre conscience de leur propre activité et de s'autoréguler tout au long de leur parcours d'apprentissage.

Les travaux de [71] abordent, par exemple, l'ajout de composants au sein des plateformes d'apprentissage pour caractériser (via des algorithmes de *machine learning*) les apprenants au regard de leurs activités et de leur préconiser ainsi des ressources pertinentes pour améliorer leur compréhension de la discipline étudiée (pour laquelle on dispose d'une carte

³ EU Sixth Framework program priority 2, Information society technology, Network of Exc.

⁴ Interaction & Collaboration Analysis supporting Teachers' and Students' Self-regulation, Jointly Executed Integrated Research Project (JEIRP)

conceptuelle liant les principaux concepts de la discipline). D'autres auteurs, comme [72], utilisent des techniques issues du web sémantique pour modéliser l'activité et les traces en lien avec des ontologies spécifiques à la discipline.

Citons également les travaux de [73] qui abordent, de leur côté, l'utilisation des fichiers de log (traces de navigation dans la plateforme) et montrent leur intérêt pour l'autorégulation des apprenants.

Ces mécanismes de collecte et de calcul sont étroitement liés aux *Learning Management Systems* (LMS) qu'ils viennent compléter et enrichir, d'autant plus que les LMS sont généralement limités en termes d'outils facilitant la collaboration [74]. Dans un environnement où ces LMS sont eux-mêmes en constante évolution [75], les plateformes d'analyse et de génération d'indicateurs doivent aussi évoluer pour tenir compte des évolutions des technologies, des usages et des dispositifs pédagogiques mis en œuvre.

Les plateformes de générateur d'indicateurs sont nombreuses et souvent dédiées à un contexte d'apprentissage spécifique.

La question de l'analyse des discussions est particulièrement importante dans le contexte du projet PASTEL. De nombreux travaux portent sur cette analyse sur la base des échanges présents dans les espaces collaboratifs, dans les espaces de chat, les forums.... Ces échanges peuvent rapidement devenir conséquents et difficiles à analyser tant pour les tuteurs que pour les apprenants eux-mêmes, nécessitant d'en synthétiser l'information pertinente et de la représenter sous une forme efficace [76], [77]. Dans [78], il s'agit en particulier d'aider à modérer des espaces synchrones de discussion en ligne (sur la base d'un mécanisme de monitoring en ligne couplé à un module off-line basé sur du *machine learning*) tandis que pour [79] il s'agit d'analyser la structure du discours pour faciliter la coordination du groupe.

Au-delà de la visualisation des interactions, il peut également être pertinent de chercher à évaluer le travail des apprenants au sein des forums et autres espaces de discussion [80].

Grâce à ces plateformes existantes, nous pouvons constituer une base d'indicateurs susceptibles d'être exploités directement dans un cadre de la TA en temps réel pour l'instrumentation de situations d'apprentissage. Néanmoins, il reste un travail important dans la conception des nouveaux indicateurs ayant pour but de qualifier et quantifier les activités des différents acteurs d'une situation d'apprentissage.

2. Bibliographie

- [1] R. Shadieff, W.-Y. Hwang, N.-S. Chen et Y.-M. Huang, «Review of Speech-to-Text Recognition Technology for Enhancing Learning,» *Educational Technology & Society*, vol. 17 (4), pp. 65-84, 2014.
- [2] Net4Voice, «WP2 - Comprehensive Learning Methodology».
- [3] M. Wald et K. Bain, «Universal access to communication and learning: the role of automatic speech recognition,» *Universal Access in the Information Society*, vol. 6 (4), pp. 435-447, 2008.
- [4] R. Deshpande, T. Tuna, J. Subhlok et L. Barker, «A Crowdsourcing Caption Editor for Educational Videos,» chez *IEEE Frontiers in Education Conference*, Madrid, Spain, 2014.
- [5] X. Che, S. Luo, H. Yang et C. Meinel, «Automatic Lecture Subtitle Generation and How It Helps,» chez *IEEE 17th International Conference on Advanced Learning Technologies*, Timisoara, Romania, 2017.
- [6] Ashwini B V et Laxmi B Ranavare, «Enhancement of Learning using Speech Recognition and Lecture Transcription: a survey,» chez *International Conference on Information and Communication Technologies*, Kochi, India, 2014.
- [7] S. Bennett, J. Hewitt, B. Mellor et C. Lyon, «Critical Success Factors for Automatic Speech Recognition in the Classroom,» chez *Universal Access in Human-Computer Interaction. Applications and Services*, Beijing, 2007.
- [8] K. Ryba et T. McIvor, «Liberated Learning: Analysis of University Students' Perceptions and Experiences with Continuous Automated Speech Recognition,» *e-Journal of Instructional Science and Technology*, vol. 9 (1), 2006.
- [9] W.-Y. Hwang, R. Shadieff, T. C. Kuo et N.-S. Chen, «Effects of Speech-to-Text Recognition Application on Learning Performance in Synchronous Cyber Classrooms,» *Education Technology & Society*, vol. 15 (1), pp. 367-380, 2012.
- [10] R. Crétin-Pirolli, F. Pirolli et V. Bettenfeld, «Projet PASTEL - Analyse des besoins,» 2017.
- [11] Y.-M. Huang, C.-J. Liu, R. Shadieff, M.-H. Shen et W.-Y. Hwang, «Investigating an application of speech-to-text recognition: a study on visual attention and learning behaviour,» *Journal of Computer Assisted Learning*, vol. 31, pp. 529-545, 2015.
- [12] N. J. Uke et R. C. Thool, «Segmentation and Organization of Lecture Video Based On Visual Contents,» *International Journal of e-Education, e-Business, e-Management and e-Learning*, vol. 2 (2), pp. 132-134, 2012.

- [13] C.-W. Ngo, F. Wang et T.-C. Pong, «Structuring Lecture Videos for Distance Learning Applications,» chez *IEEE Fifth International Symposium on Multimedia Software Engineering*, 2003.
- [14] T. Tuna, M. Joshi, V. Varghese, R. Deshpande, J. Subhlok et R. Verma, «Topic Based Segmentation of Classroom Videos,» chez *IEEE Frontiers in Education Conference (FIE)*, El Paso, Texas, 2015.
- [15] G. Alharbi, R. W. M. Ng et T. Hain, «Annotating Meta-discourse in Academic Lectures from Different Disciplines,» chez *SLaTE*, Leipzig, 2015.
- [16] O. Aubert et Y. Prié, «Annotations, a key asset for video-based e-learning,» chez *Internet of Education*, Ljubljana, 2013.
- [17] L. Elliot, S. Foster et M. Stinson, «Student Study Habits Using Notes from a Speech-to-Text Support Service,» *Exceptional Children*, vol. 69 (1), pp. 25-40, 2002.
- [18] M. Papadopoulos et E. Pearson, «Improving the Editing Process of Automatically Produced Lecture Transcripts Based on Natural Language Analysis,» chez *IEEE International Conference on Intelligent Computing and Intelligent Systems*, 2010.
- [19] R. Shadiev, W.-Y. Hwang et Y.-M. Huang, «Investigating Learning Strategies of Using Texts Generated by Speech to Text Recognition Technology in Traditional Classroom,» chez *AECT International Conference on the Frontier in e-Learning Research*, 2013.
- [20] R. Shadiev, W.-Y. Hwang et Y.-M. Huang, «Investigating applications of Speech-to-Text recognition to assist learning in online and traditional classrooms,» *International Journal of Humanities and Arts Computing*, vol. 8, pp. 179-189, 2014.
- [21] L. S. Kate, M. M. Waghmare et A. Priyadarshi, «An approach for automated video indexing and video search in large lecture video archives,» chez *International Conference on Pervasive Computing (ICPC)*, 2015.
- [22] S.-W. Li et V. Zue, «Linking MOOC Courseware to Accomodate Diverse Learner Backgrounds,» chez *SLaTE 2015*, Leipzig, 2015.
- [23] F. Gasparetti, C. De Medio, C. Limongelli, F. Sciarrone et M. Temperini, «Prerequisites between learning objects: Automatic extraction based on a machine learning approach,» *Telematics and Informatics*, p. in press, 2017.
- [24] A. Weggerle, P. Schmidt et P. Schulthess, «Speech to Multi-Media document transcription for university lectures,» chez *ICERI2009 Conference*, Madrid, Spain, 2009.
- [25] M. Wald, «An exploration of the potential of Automatic Speech Recognition to assist and enable receptive communication in higher education,» *ALT-J, Research in Learning Technology*, vol. 14 (1), pp. 9-20, 2006.

- [26] P. Cerva, J. Silovsky, J. Zdansky, O. Smola, K. Blavka, K. Palecek, J. Nouza et J. Malek, «Browsing, Indexing and Automatic Transcription of Lectures for Distance Learning,» chez *IEEE 14th International Workshop on Multimedia Signal Processing (MMSP)*, 2012.
- [27] O. Aubert, Y. Prié et C. Canellas, «Leveraging video annotations in video-based e-learning,» chez *7th International Conference on Computer Supported Education (CSEU)*, Barcelone, Spain, 2014.
- [28] T. S. Hasselbring et C. H. Williams Glaser, «Use of Computer Technology to Help Students with Special Needs,» *Children and Computer Technology*, vol. 10 (2), pp. 102-122, 2000.
- [29] M. Wald, «Using Automatic Speech Recognition to Enhance Education for All Students: Turning a Vision into Reality,» chez *ASEE/IEEE Frontiers in Education Conference*, Savannah, GA, 2004.
- [30] E. A. O'Hare et M. F. McTear, «Speech recognition in the secondary school classroom: an exploratory study,» *Computers & Education*, vol. 33, pp. 27-45, 1999.
- [31] B. Z. Halimah, A. Azlina, P. Behrang et W. O. Choo, «Voice Recognition System for the Visually Impaired: Virtual Cognitive Approach,» chez *International Symposium on Information Technology*, 2008.
- [32] M. R. Mirzaei, S. Ghorshi et M. Mortazavi, «Audio-visual speech recognition techniques in augmented reality environments,» *The Visual Computer*, vol. 30 (3), pp. 245-257, 2014.
- [33] A. Chern, Y.-H. Lai, Y.-P. Chang, Y. Tsao, R. Y. Chang et H.-W. Chang, «A Smartphone-Based Multi-Functional Hearing Assistive System to Facilitate Speech Recognition in the Classroom,» *IEEE Access*, vol. 5, pp. 10339-10351, 2017.
- [34] N. Heni, «Design of Emotional Educational System Mobile Games for Autistic Children,» chez *International Conference on Advanced Technologies for Signal and Image Processing - ATSIP'2016*, Monastir, Tunisia, 2016.
- [35] M. Ramirez, M. Sotaquira, A. De La Cruz, E. Maria, G. Avellaneda et A. Ochoa, «An Automatic Speech Recognition System for Helping Visually Impaired Children to Learn Braille,» chez *XXI Symposium on Signal Processing, Images and Artificial Vision (STSIVA)*, 2016.
- [36] E. McAteer, T. Anderson, M. Orr, A. Demissie et E. Woherem, «Computer-assisted learning and groupwork: the design of an evaluation,» *Computers Educ.*, vol. 17 (1), pp. 41-47, 1991.
- [37] R. Shadiev, Y.-M. Huang, W.-Y. Hwang et N. Shadiev, «Investigating the effectiveness of speech-to-text recognition application on learning performance in traditional learning environment,» chez *International Conference on Advanced Learning Technologies - ICALT*, 2015.

- [38] Y.-M. Huang, R. Shadiev et W.-Y. Hwang, «Investigating the effectiveness of speech-to-text recognition applications on learning performance and cognitive load,» *Computers & Education*, vol. 101, pp. 15-28, 2016.
- [39] R. Shadiev, Y.-M. Huang et T.-T. Wu, «Facilitating comprehension of non-native English speakers during lectures in English with STR-texts,» chez *International Conference on Advanced Learning Technologies - ICALT*, 2016.
- [40] R. Shadiev, Y.-M. Huang, W.-Y. Hwang et N. Shadiev, «Investigating visual attention of students with different learning ability on texts generated by speech-to-text recognition,» chez *International Conference on Advanced Learning Technologies - ICALT*, 2014.
- [41] B. Perelmutter, K. K. McGregor et K. R. Gordon, «Assistive technology interventions for adolescents and adults with learning disabilities: An evidence-based systematic review and meta-analysis,» *Computers & Education*, vol. 114, pp. 139-163, 2017.
- [42] W. Goddard, L. Kaplan, J. Kuehnle et M. Beglau, «Voice recognition and Speech-to-Text pilot implementation in primary general education technology-rich eMINTS classrooms,» 2007.
- [43] J. Jiménez Dorado, P. Revuelta Sanz, L. Moreno Lopez et A. Iglesias Maqueda, «Evaluating the Use of Speech Technologies In the Classroom: The APEINTA Project,» chez *World Conference on Educational Multimedia, Hypermedia & Telecommunications, ED-MEDIA 2010*, 2010.
- [44] S. M. McCrocklin, «Pronunciation learner autonomy: The potential of Automatic Speech Recognition,» *System*, vol. 57, pp. 25-42, 2016.
- [45] D. Coniam, «Voice recognition software accuracy with second language speakers of English,» *System*, vol. 27, pp. 49-64, 1999.
- [46] L. Indrayanti, T. Usagawa, Y. Chisaki et T. Dutono, «Evaluation of Pronunciation by means of Automatic Speech Recognition System for Computer Aided Indonesian Language Learning,» chez *International Conference on Information Technology Based Higher Education and Training*, 2006.
- [47] S. M. McCrocklin, «Pronunciation learner autonomy: The potential of Automatic Speech Recognition,» *System*, vol. 57, pp. 25-42, 2016.
- [48] L. Czap, J. M. Pintér et A. K. Varga, «Segmentation the Speech of Hard of Hearing Children,» chez *18th International Carpathian Control Conference (ICCC)*, 2017.
- [49] J. H. Hansen et T. Hasan, «Speaker Recognition by Machines and Humans - A tutorial review,» *IEEE Signal Processing Magazine*, vol. 74, pp. 74-99, Nov 2015.

- [50] W. Dong, B. Lepri, F. Pianesi et A. Pentland, «Modeling Functional Roles Dynamics in Small Group Interactions,» *IEEE Transactions on Multimedia*, vol. 15 (1), pp. 83-95, Jan 2013.
- [51] H. Salamin, S. Favre et A. Vinciarelli, «Automatic Role Recognition in Multiparty Recordings: Using Social Affiliation Networks for Feature Extraction,» *IEEE Transactions on Multimedia*, vol. 11 (7), pp. 1373-1380, Nov 2009.
- [52] N. P. Garg, S. Favre, H. Salamin, D. Hakkani Tür et A. Vinciarelli, «Role Recognition for Meeting Participants: an Approach Based on Lexical Information and Social Network Analysis,» chez *ACM International Conference on Multimedia*, Vancouver, Canada, 2008.
- [53] D. Sanchez-Cortes, O. Aran, M. Schmid Mast et D. Gatica-Perez, «A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups,» *IEEE Transactions on Multimedia*, vol. 14 (3), pp. 816-832, Juin 2012.
- [54] A. Sapru et H. Bourlard, «Detecting speaker roles and topic changes in multiparty conversations using latent topic models,» chez *INTERSPEECH 2014*, Singapore, 2014.
- [55] T. C. T. Kuo, R. Shadiev, W.-Y. Hwang et N.-S. Chen, «Effects of Applying STR for Group Learning Activities on Learning Performance in a Synchronous Cyber Classroom,» chez *IEEE International Conference on Advanced Learning Technologies - ICALT*, 2011.
- [56] T. C. T. Kuo, R. Shadiev, W.-Y. Hwang et N.-S. Chen, «Effects of applying STR for group learning activities on learning performance in a synchronous cyber classroom,» *Computers & Education*, vol. 58, pp. 600-608, 2012.
- [57] R. Shadiev et Y.-M. Huang, «Facilitating cross-cultural understanding with learning activities supported by speech-to-text recognition and computer-aided translation,» *Computers & Education*, vol. 98, pp. 130-141, 2016.
- [58] R. Ranchal, T. Taber-Doughty, Y. Guo, K. Bain, H. Martin, J. P. Robinson et B. S. Duerstock, «Using Speech Recognition for Real-Time Captioning and Lecture Transcription in the Classroom,» *IEEE Transactions on Learning Technologies*, vol. 6 (4), pp. 299-311, Dec 2013.
- [59] M. Wald, «Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time,» chez *ICCHP 2006*, 2006.
- [60] M. Wald et K. Bain, «Enhancing the Usability of Real-Time Speech Recognition Captioning Through Personalised Displays and Real-Time Multiple Speaker Editing and Annotation,» chez *Universal Access in HCI, Part III, HCII 2007*, 2007.
- [61] Y. Akita, N. Kuwahara et T. Kawahara, «Automatic Classification of Usability of ASR Result for Real-time Captioning of Lectures,» chez *Proceedings of APSIPA Annual Summit and Conference 2015*, 2015.

- [62] A. Shiv Sharma et R. Bhalley, «ASR – A real-time speech recognition on portable devices,» chez *International Conference on Advances in Computing, Communication, & Automation (ICACCA)*, 2016.
- [63] A. Corbi et D. Burgos, «Review of Current Student-Monitoring Techniques used in eLearning-Focused recommender Systems and Learning analytics. The Experience API & LIME model Case Study,» *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 2, n° %17, pp. 44-52, 09 2014.
- [64] R. Mazza et L. Botturi, «Monitoring an Online Course with the GISMO Tool: A Case Study,» *International Journal of Interactive Learning Research*, vol. 18, pp. 251-265, 2007.
- [65] D. A. Gómez-Aguilar, Á. Hernández-García, F. J. García-Peñalvo et R. Therón, «Tap into visual analysis of customization of grouping of activities in eLearning,» *Computers in Human Behavior*, vol. 47, pp. 60-67, 2015.
- [66] P. Jermann, A. Soller et M. Muehlenbrock, «From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning,» chez *Proceedings of the First European Conference on Computer-Supported Collaborative Learning*, Maastricht, The Netherlands, 2001.
- [67] A. Alowayr et A. Badii, «Review of monitoring tools for e-learning platforms,» *International Journal of Computer Science & Information Technology*, vol. 6, n° %13, pp. 79-86, 2014.
- [68] A. Dimitracopoulou, A. Martinez Mones, Y. Dimitriadis, A. Morch, S. Ludvigsen, A. Herrer, U. Hoppe, B. Barros, F. Verdejo, C. Hulshof, T. Jong, G. Fessakis, A. Petrou, K. Lund, M. Baker, P. Jermann, P. Dillenbourg, V. Kollias et Vosniadou, «State of the art on interaction analysis: Interaction analysis indicators.,» 2004.
- [69] A. Hernández-García, I. González-González, A. I. Jiménez-Zarco et J. Chaparro-Peláez, «Applying social learning analytics to message boards in online distance learning: A case study,» *Computers in Human Behavior*, vol. 47, pp. 68-80, 2015.
- [70] W. Xing, R. Guo, E. Petakovic et S. Goggins, «Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory,» *Computers in Human Behavior*, vol. 47, pp. 168-181, 2015.
- [71] D. D. Burdescu et M. C. Mihăescu, «Building Intelligent E-Learning Systems by Activity Monitoring and Analysis,» chez *Multimedia Services in Intelligent Environments. Smart Innovation, Systems and Technologies*, vol. 3, Springer, Berlin, Heidelberg, 2010, p. 153–174.



Laboratoire d'Informatique de l'Université du Maine

- [72] S. Softic, B. Taraghi, M. Ebner, L. De Vocht, E. Mannens et R. Van de Walle, «Monitoring Learning Activities in PLE Using Semantic Modelling of Learner Behaviour,» *Human Factors in Computing and Informatics*, pp. 74-90, 2013.
- [73] D. Jeske, J. Backhaus et C. Stamov Roßnagel, «Self-regulation during e-learning: using behavioural evidence from navigation log files,» *Journal of Computer Assisted Learning*, vol. 30, pp. 272-284, 2014.
- [74] C. Brooks, R. Panesar et J. Greer, «Awareness and Collaboration in the iHelp Courses Content Management System,» chez *Innovative Approaches for Learning and Knowledge Sharing*, 2006.
- [75] S. Mehra, «Emerging Trends in Learning Management Systems,» *International Journal of Computer Science Issues*, vol. 12, n° 11 (2), pp. 79-83, 2015.
- [76] J. Donath, «A semantic approach to visualizing online conversations,» *Communications of the ACM*, vol. 45, pp. 45-49, 2002.
- [77] W. Gibbs, V. Olexa et R. Bernas, «A Visualization Tool for Managing and Studying Online Communications,» *International Journal of Educational Technology and Society*, vol. 9, pp. 232-243, 2006.
- [78] R. de Groot, R. Drachman, R. Hever, B. Schwarz, U. Hoppe, A. Harrer, M. De Laat, R. Wegerif, B. McLaren et B. Baurens, «Computer Supported Moderation of E-Discussions: the ARGUNAUT Approach,» chez *International Conference of Computer Supported Collaborative Learning*, New Jersey, USA, 2007.
- [79] M. Gerosa, M. Pimentel, H. Fuks et G. Lucena, «Analyzing Discourse Structure to Coordinate Educational Forums,» chez *The 7th International Conference on Intelligent Tutoring Systems*, Maceiò, Alagoas, Brazil, 2004.
- [80] M. Shaul, «Assessing online discussion forum participation,» *International Journal of Information and Communication Technology Education*, vol. 3, pp. 39-46, 2007.