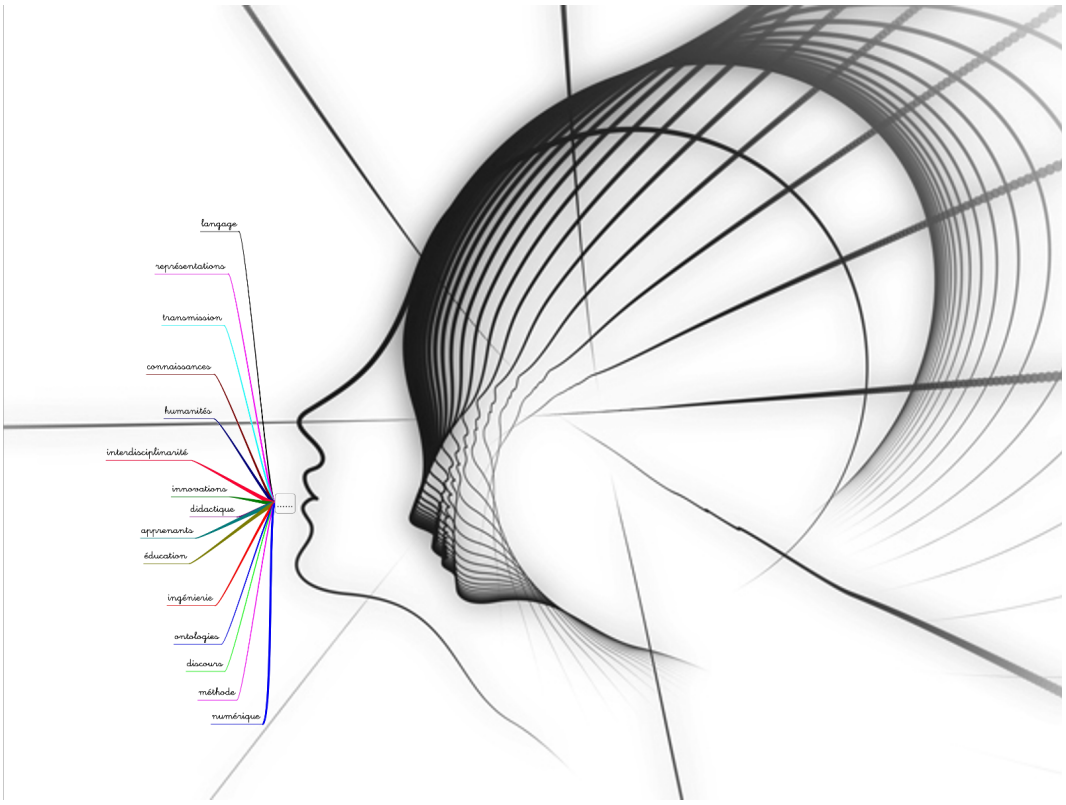


Colloque Jeunes Chercheurs 2019 PRAXILING
Représentations et transmission des connaissances à la lumière de l'innovation numérique

Actes

-

7-8 Novembre 2019

Table des matières

Bienvenue sur le site des actes du colloque **5**

Introduction **7**
Axes de réflexion 7
Comité scientifique 11
Comité d'organisation 11

1 Axe 1 : Constitution et analyse de corpus textuels et oraux, les outils et leurs pratiques 13
1.1 Productions participatives de corpus annotés : des modèles encore incertains 13
1.2 Prediction of Missing Semantic Relations in Lexical-Semantic Network using Random Forest Classifier 56
1.3 La logique hyperdocumentaire dans la structuration de l'information et sa visualisation . 66
1.4 Démarche de recueil de pratiques en vue de l'émergence d'un modèle d'élaboration didactique d'un cours de thermodynamique dans l'enseignement supérieur. 74
1.5 Annotation et exploitation numérique d'un corpus d'écrits scolaires corrigés en France et en Italie 119
1.6 Vers une résolution des relations anaphoriques dans la communication électronique médiée 138
1.7 La constitution d'un corpus d'ellipses erronées pour la représentation et la transmission des connaissances 185
1.8 Aspects acoustique et articulatoire du bégaiement dans un contexte bilingue : cas des locuteurs mina-français 213

2 Axe 2 : Numérique, cognition, apprentissages et représentations psychosociales 261
2.1 Le rôle du prof. dans les dispositifs numériques : une vision diachronique 261
2.2 MOOC : lorsque le Maître se met en scène 262
2.3 L'intégration, l'usage et l'apport de la compétence technologique dans l'enseignement-apprentissage des élèves du primaire : quels jalons et quels enjeux 282
2.4 Enquête sur la littératie numérique des étudiants chinois pour l'apprentissage du français 282
2.5 Des doctorant-e-s SHS face à l'innovation numérique : enjeux techniques, politique, pédagogiques. 292
2.6 La confiance et son impact dans le domaine du numérique 299

3 Axe 3 : Impacts sociétaux des nouveaux processus d'acquisition : construction, co-construction et modélisation des savoirs. 313
3.1 Quelle maîtrise dans les milieux de savoir du digital ? 313
3.2 Etude sur l'Argumentation de Collégiens au Sujet d'une Controverse Socio-scientifique Présentée en Vidéo 329
3.3 Le numérique au service de la transmission des savoirs. 335
3.4 Instrumentalisation de l'intelligence collective ? 347

4 Conclusions du colloque **357**

Bienvenue sur le site des actes du colloque

Introduction

Les projets relevant des Humanités Numériques permettent de dépasser le cloisonnement disciplinaire des acteurs de la recherche et de l'apprentissage. La capacité innovatrice devient un facteur qui conditionne l'évolution du domaine des sciences humaines et sociales. Les apports théoriques, et méthodologiques issus du croisement des recherches en linguistique, en management des connaissances, en traitement du langage naturel, en sciences cognitives, en apprentissage profond, en fouille de données (pour ne citer que ceux-ci) et leurs applications directes dans la société témoignent de l'extrême vigueur de ce domaine. Le numérique n'est pas seulement un ensemble de techniques qui permettent aux organisations de travailler plus efficacement et de développer de nouveaux canaux de relation avec les usagers : il constitue aujourd'hui le principal levier de transformation des organisations apprenantes. Le numérique transforme nos représentations en particulier dans le cadre des actions d'intermédiation entre enseignants et apprenants. Co-innovation, co-création, standards, plateformes permettent aux projets de recherche de devenir des "écosystèmes" au sens de réseaux de dépendances et d'échanges. En leur sein, des acteurs qui appartiennent à des disciplines différentes collaborent dans des configurations qui sont amenées à évoluer en permanence dans leurs protocoles méthodologiques.

Axes de réflexion

Axe 1 : Constitution et analyse de corpus textuels et oraux, les outils et leurs pratiques.

Les imaginaires des pratiques d'enseignement liées au numérique, incluant pour l'enseignant la crainte que l'utilisation du numérique ne soit synonyme de substitution (fantasmes liés au tuteur-machine ou au tuteur automatisé) nécessitent que l'ensemble des outils et des pratiques soient explicitées et accompagnées (traitement automatique de la langue naturelle écrite et orale pour la recherche d'informations, la détection de relation, l'analyse de sentiments, ou la fouille d'arguments).

Du concept à la connaissance :

- pratiques d'extraction et d'acquisition de connaissances didactiques ou pédagogiques, fouille de données,
- peuplement d'ontologies ou de réseaux lexicaux, annotation sémantique,
- standardisation des annotations et du balisage,
- ingénierie de systèmes collaboratifs,
- acquisition de connaissances à partir de données non structurées, à partir d'interactions conversationnelles,
- traitement et raisonnement sur les connaissances.

Quelles nouvelles pistes pourrait-on envisager face aux besoins des enseignants et des apprenants dans le cadre de l'évolution des technologies ?

Bibliographie Axe 1

BERTOLINI, Marco, 2014. Formation innovante : BYOD, hybride, netware, cloud, deep learning. Formation 3.0 [en ligne]. 29 octobre 2014. [Consulté le 24 janvier 2019]. Disponible à l'adresse : https://format30.com/2014/10/29/formation-innovante-byod-hybride-netware-cloud-deep-learning/

CAPPELLINI, Marco, 2018. Synthèse et discussion de Numérique et formation des enseignants de langue–Pistes et imaginaires. Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication. 2018.

CIEKANSKI, Maud, 2014. Accompagner l'apprentissage des langues à l'heure du numérique–Évolution des problématiques et diversité des pratiques. Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication. 2014. Vol. 17.

CORMERAIS, Franck, LE DEUFF, Olivier, LAKEL, Amar et PUCHEU, David, 2016. Les SIC à l'épreuve du digital et des Humanités : des origines, des concepts, des méthodes et des outils. Revue française des sciences de l'information et de la communication [en ligne]. 1 janvier 2016. N° 8. [Consulté le 4 février 2019]. DOI 10.4000/rfsic.1820. Disponible à l'adresse : http://journals.openedition.org/rfsic/1820

FORT, Karën, ADDA, Gilles et COHEN, K. Bretonnel, 2016. Éthique et traitement automatique des langues et de la parole : entre truismes et tabous. Traitement Automatique des Langues. 2016. Vol. 57, n° 2, pp. 7-19. Inventaire d'outils TIC et de ressources éducatives ouvertes, [sans date]. European Centre for Modern Languages of the Council of Europe [en ligne]. [Consulté le 11 mars 2019]. Disponible à l'adresse : https://www.ecml.at/ECML-Programme/Programme2012-2015/ICT-REVandmoreDOTS/ICT/tabid/1906/Default.aspx

MINICHIELLO, Federica, 2017. Machines, données et apprentissage : relations et enjeux. Revue internationale d'éducation de Sèvres [en ligne]. 1 avril 2017. N° 74, pp. 12-15. [Consulté le 24 janvier 2019]. DOI 10.4000/ries.5755. Disponible à l'adresse : http://journals.openedition.org/ries/5755

PÉLISSIER, Chrysta, 2017. Accompagner le chercheur en SHS à l'ère des humanités numériques. Des outils pour le développement d'une activité cognitive augmentée. Les Cahiers du numérique [en ligne]. 2017. Vol. 13, n° 3-4, pp. 167-194. Disponible à l'adresse : https://www.cairn.info/revue-les-cahiers-du-numerique-2017-3-4-page-167.htm Cairn.info

TRAUSAN-MATU, Stefan, DASCALU, Mihai et DESSUS, Philippe, 2012. Textual Complexity and Discourse Structure in Computer-Supported Collaborative Learning. In : Intelligent Tutoring Systems [en ligne]. Berlin, Heidelberg : Springer Berlin Heidelberg. pp. 352-357. [Consulté le 9 février 2019]. ISBN 978-3-642-30949-6. Disponible à l'adresse : http://link.springer.com/10.1007/978-3-642-30950-2_46

Axe 2 : Numérique, cognition, apprentissages et représentations psychosociales

La place et la fonction de l'enseignant évoluent constamment, au fil des innovations numériques et recommandations officielles. Qu'en est-il des représentations psychosociales ? Que ce soit au niveau individuel ou au niveau collectif, on peut distinguer quatre problématiques :

- le niveau des exigences auxquelles faire face (sentiment d'urgence, répétitivité des tâches, volume et temporalité des sollicitations, asynchronicité des actions, multiplicité des agents sollicitant, complexité des procédures, ergonomie des outils)
- le degré d'autonomie et l'équilibre "vie privée"/"vie professionnelle", (pluralité des acteurs et des avis, problème d'organisation, temps de réponse aux sollicitations que l'on émet,…) #charge de travail
- le niveau de soutien (des différents acteurs de l'apprentissage),

— le besoin de reconnaissance (sens donné au travail produit, représentation du collectif, prise en compte des contributions, évaluation/pertinence des autres acteurs faisant partie du réseau, estime de soi, valorisation) #co-working #personal-branding

La validité et la légitimité des savoirs transmis via le numérique est-elle remise en question ?

Quel est l'impact de l'utilisation du numérique sur le positionnement de l'enseignant en tant qu'expert ? Quelle est sa place dans cette nouvelle configuration de l'enseignement (gestion des trois notions : distance, absence, présence) ?

Reste-t-il l'expert de ses savoirs ? L'innovation numérique est telle qu'il doit s'y adapter afin de répondre aux nouveaux besoins et aux nouvelles compétences de ses élèves.

Bibliographie Axe 2

CAPPELLINI, Marco, 2018. Synthèse et discussion de Numérique et formation des enseignants de langue–Pistes et imaginaires. Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication. 2018.

DEVELOTTE, Christine, 2010. Réflexions sur les changements induits par le numérique dans l'enseignement et l'apprentissage des langues. Ela. Études de linguistique appliquée. 2010. N° 4, pp. 445-464.

JÉZÉGOU, Annie, 2010. Créer de la présence à distance en e-learning. Cadre théorique, définition, et dimensions clés. Distances et savoirs [en ligne]. 30 juin 2010. Vol. 8, n° 2, pp. 257-274. [Consulté le 16 janvier 2019]. DOI 10.3166/ds.8.257-274. Disponible à l'adresse : http://ds.revuesonline.com/article.jsp?articleId=15238

OLLIVIER, Christian, 2016. Gestion du pouvoir dans la co-construction de savoirs . Le rôle des contributions péritextuelles contextualisées dans la coélaboration d'objets de savoir partagés. Distances et médiations des savoirs. Distance and Mediation of Knowledge [en ligne]. 29 septembre 2016. N° 15. [Consulté le 20 janvier 2019]. DOI 10.4000/dms.1558. Disponible à l'adresse : http://journals.openedition.org/dms/1558

TRINQUIER, Marie-Pierre, 2009. Approche psychosociale des représentations et des pratiques d'enseignement…. à propos de l'utilisation de l'ordinateur à l'école. Travail et formation en éducation. 2009. N° 3.

Axe 3 : Impacts sociétaux des nouveaux processus d'acquisition : construction, co-construction et modélisation des savoirs.

La constante et rapide évolution des nouvelles technologies permet l'émergence de nouvelles formes d'apprentissages. Mais qu'en est-il de leur légitimité au sein de la société ? Quel impact ont-elles sur les représentations de l'apprentissage ? Comment l'apprenant perçoit-il la construction et la co-construction des savoirs et quel impact ses représentations ont sur ses propres apprentissages ?

La formation proposée a-t-elle la même valeur si elle est dispensée à distance ? Si les apprenants évoluent à leur rythme, la valeur de la formation est-elle à leurs yeux, essentielle ? Suivre une formation à distance est-elle valorisante au regard de la société ? Un diplôme obtenu à distance a-t-il la même valeur sur le marché du travail ? Peuvent-ils réellement se prévaloir des mêmes compétences validées par leurs camarades en présentiel ?

Quels sont les incidences et les conséquences du numérique sur les processus d'apprentissage et en didactique :

— ubiquité de l'apprentissage (Ubiquitous Learning UL),
— apprenant mobile (mobile learning),
— jeux sérieux (Game-Based Learning),
— réseaux sociaux.

Bibliographie Axe 3

BERTOLINI, Marco, 2014. Formation innovante : BYOD, hybride, netware, cloud, deep learning. Formation 3.0 [en ligne]. 29 octobre 2014. [Consulté le 24 janvier 2019]. Disponible à l'adresse : https://format30.com/2014/10/29/formation-innovante-byod-hybride-netware-cloud-deep-learning/

BOCQUET, François, BRUILLARD, Éric, CORNU, Bernard, GUIGNOLET, Joël, MOATTI, Daniel et VÉRAN, Jean-Pierre, 2014. Le numérique à l'école : évolution ou révolution pédagogique ?. Un état des réflexions en France - Table ronde. Revue internationale d'éducation de Sèvres [en ligne]. 14 décembre 2014. N° 67, pp. 103-118. [Consulté le 4 février 2019]. DOI 10.4000/ries.4129. Disponible à l'adresse : http://journals.openedition.org/ries/4129

BRUDERMANN, Cédric, 2018. Interagir avec la « machine » à l'ère du numérique : des représentations des étudiants à l'évaluation de leur performance académique en anglais. Revue internationale de pédagogie de l'enseignement supérieur [en ligne]. 26 mars 2018. Vol. 34, n° 34-1. [Consulté le 16 janvier 2019]. Disponible à l'adresse : http://journals.openedition.org/ripes/1278

DEGACHE, Christian et MANGENOT, François, 2007. 36 | 2007 Échanges exolingues via internet et appropriation des langues-cultures. [en ligne]. 2007. [Consulté le 16 janvier 2019]. Disponible à l'adresse : http://journals.openedition.org/lidil/2333

DERVIN, Fred et VLAD, Monica, 2010. Pour une cyberanthropologie de la communication interculturelle – Analyse d'interactions en ligne entre étudiants finlandais et roumains. Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication [en ligne]. 28 février 2010. N° Vol. 13. [Consulté le 16 janvier 2019]. DOI 10.4000/alsic.1399. Disponible à l'adresse : http://journals.openedition.org/alsic/1399

DEVELOTTE, Christine, 2010. Réflexions sur les changements induits par le numérique dans l'enseignement et l'apprentissage des langues. Ela. Etudes de linguistique appliquee [en ligne]. 2010. Vol. n°160, n° 4, pp. 445-464. [Consulté le 16 janvier 2019]. Disponible à l'adresse : https://www.cairn.info/revue-ela-2010-4-page-445.htm

DOWNES, Stephen, 2007. Learning networks in practice. . 2007.

GOULMOT, Isabelle, 2016. Mooc, Cooc, Spoc, Sooc… Quelles différences ? Journal du Net [en ligne]. 12 juin 2016. [Consulté le 4 février 2019]. Disponible à l'adresse : https://www.journaldunet.com/management/expert/64562/mooc--cooc--spoc--sooc----quelles-differences.shtml

JÉZÉGOU, Annie, 2010. Créer de la présence à distance en e-learning. Cadre théorique, définition, et dimensions clés. Distances et savoirs [en ligne]. 30 juin 2010. Vol. 8, n° 2, pp. 257-274. [Consulté le 16 janvier 2019]. DOI 10.3166/ds.8.257-274. Disponible à l'adresse : http://ds.revuesonline.com/article.jsp?articleId=15238

MINICHIELLO, Federica, 2017. Machines, données et apprentissage : relations et enjeux. Revue internationale d'éducation de Sèvres [en ligne]. 1 avril 2017. N° 74, pp. 12-15. [Consulté le 24 janvier 2019]. DOI 10.4000/ries.5755. Disponible à l'adresse : http://journals.openedition.org/ries/5755

OLLIVIER, Christian, 2016. Gestion du pouvoir dans la co-construction de savoirs . Le rôle des contributions péritextuelles contextualisées dans la coélaboration d'objets de savoir partagés. Distances et médiations des savoirs. Distance and Mediation of Knowledge [en ligne]. 29 septembre 2016. N° 15. [Consulté le 20 janvier 2019]. DOI 10.4000/dms.1558. Disponible à l'adresse : http://journals.openedition.org/dms/1558

PERAYA, Daniel, 2018. Technologies, innovation et niveaux de changement : les technologies peuvent-elles modifier la forme universitaire ? Distances et médiations des savoirs. Distance and Mediation of Knowledge. 2018. Vol. 2018, n° 21. SÉNAT, Commission de la culture, de l'éducation et de la communication, 2018. Prendre en main notre destin numérique : l'urgence de la formation [en ligne]. [Consulté le 4 février 2019]. Disponible à l'adresse : http://www.senat.fr/rap/r17-607/r17-607_mono.html

TREMION, Virginie et DERVIN, Fred, 2018. De Cultura aux MOOCs de communication interculturelle : quelles opportunités pour l'apprentissage interculturel à distance ? International Journal of E-Learning & Distance Education / Revue internationale du e-learning et la formation à distance [en ligne]. 20 mai 2018. Vol. 33, n° 1. [Consulté le 16 janvier 2019]. Disponible à l'adresse : http://www.ijede.ca/index.php/jde/article/view/1048

Comité scientifique

Alidières-Dumonceaud Lucie ITIC Praxiling UPVM3,
Auger Nathalie PRAXILING UPVM3,
Brudermann Cédric, DILTEC UPMC
Charnet Chantal, PRAXILING UPVM3
Ciekanski Maud, ATILF U. LORRAINE
Develotte Christine, ICAR ENS LYON
Dodane Christelle, PRAXILING UPVM3
Fort Karën, STIH SORBONNE U.
Frontini Francesca, PRAXILING UPVM3
Hirsch Fabrice, PRAXILING UPVM3
Jackiewicz Agata, PRAXILING UPVM3
Jacquot Clémence, PRAXILING UPVM3
LeDeuff Olivier, MICA U. BORDEAUX
Luschka Félix, PRAXILING UPVM3
Metz Mailles Viard Stéphanie, PRAXILING UPVM3
Mangenot François, LILIDEM U. GRENOBLE3
Nowakowska Aleksandra, PRAXILING UPVM3
Pelissier Chrysta, PRAXILING UPVM3
Perea François, PRAXILING UPVM3
Richard Arnaud, PRAXILING UPVM3
Sauvage Jérémi, PRAXILING UPVM3
Siouffi Gilles, STIH SORBONNE U.
Steuckardt Agnès, PRAXILING UPVM3
Trémion Virginie, ISP U. CATHOLIQUE DE PARIS
Verdier Maud, PRAXILING UPVM3
Verine Bertrand,PRAXILING UPVM3

Comité d'organisation

BIALES Anne-Laure
DALLE Laurine
DELEBARRE Justine
TOUBOL Pierre-Bernard
TOURRON Fanny

1

Axe 1 : Constitution et analyse de corpus textuels et oraux, les outils et leurs pratiques

1.1 Productions participatives de corpus annotés : des modèles encore incertains

Conférence invitée - Karën FORT, STIH, Université Paris-Sorbonnes

"La production participative (ou crowdsourcing) consiste à faire produire à une foule plus ou moins importante de participant · e · s, des votes, des avis, des données, le plus souvent via une application en ligne. Après avoir défini plus précisément les différents types de production participative, je présenterai plusieurs expériences menées dans le domaine du traitement automatique des langues (TAL) pour la production de données langagières, en particulier des corpus annotés. Outre nos (relatifs) succès, je tenterai d'analyser les limites de ce type de plateformes et de montrer qu'au-delà de l'effet de mode et de la gratuité supposée, la production participative peut, dans certaines conditions encore mal maîtrisées, avoir un réel impact scientifique et humain."

CC BY NC SA

Productions participatives de corpus annotés : des modèles encore incertains

Karën Fort

karen.fort@sorbonne-universite.fr

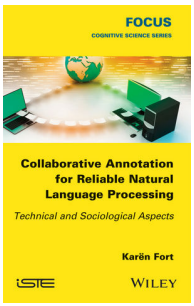
8 novembre 2019

LETTRES SORBONNE UNIVERSITÉ — Loria Laboratoire lorrain de recherche en informatique et ses applications

D'où je parle

Voir http://karenfort.org/

- Création de ressources langagières pour le Traitement automatique des langues (TAL)



- Ethique et TAL

Production participative (*crowdsourcing*)

Jeux ayant un but que j'ai participé à créer :

Portail de jeux pour les langues et atelier récurrent :

Games4NLP

Les productions participatives

Les jeux ayant un but : l'exemple de ZombiLingo

Les incertitudes du modèle

Conclusion et perspectives

Les productions participatives
Des participants qu'on ne choisit pas
Au-delà des mythes
Redéfinition

Les jeux ayant un but : l'exemple de ZombiLingo

Les incertitudes du modèle

Conclusion et perspectives

Production participative / myriadisation (*crowdsourcing*)

> **Crowdsourcing** *is "the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call."[Howe, 2006]*

- pas d'identification ou de sélection des participants *a priori* (appel ouvert)
- massif (en termes de production et de participation)
- (relativement) peu cher

Des réussites remarquables

`Wikipédia`[1] (octobre 2019) :

- plus de **158 million d'articles** en **279** langues
- plus de 800 millions de pages de la Wikipédia française vues en septembre

`Distributed Proofreaders (Gutenberg Project)`[2] :

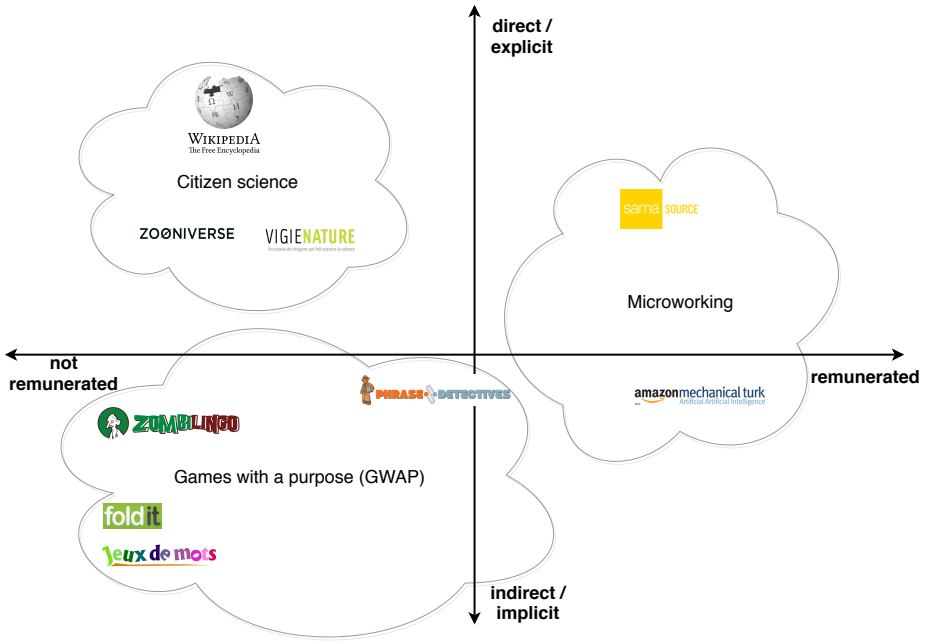
- près de **40 000** livres numérisés et corrigés

Numérisation des déclarations de conflits d'intérêts des élus[3] :

- 11 095 extraits de déclarations saisis **en moins d'une semaine**
- près de 8 000 participants

1. https://stats.wikimedia.org/v2/#/all-wikipedia-projects
2. https://www.pgdp.net/c/
3. http://regardscitoyens.org/interets-des-elus/

Les productions participatives



Les productions participatives
Des participants qu'on ne choisit pas
Au-delà des mythes
Redéfinition

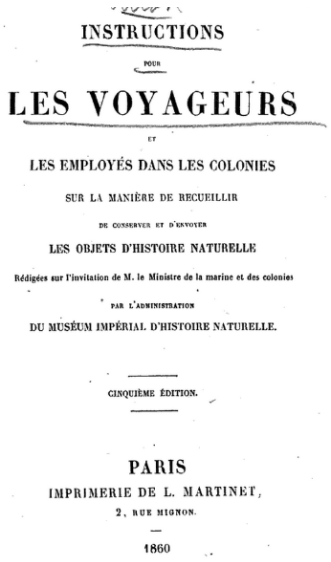
Les jeux ayant un but : l'exemple de ZombiLingo

Les incertitudes du modèle

Conclusion et perspectives

Mythe #1 : « C'est nouveau ! »

Instructions pour les voyageurs et les employés des colonies



INSTRUCTIONS

POUR

LES VOYAGEURS

ET

LES EMPLOYÉS DANS LES COLONIES

SUR LA MANIÈRE DE RECUEILLIR

DE CONSERVER ET D'ENVOYER

LES OBJETS D'HISTOIRE NATURELLE

Rédigées sur l'invitation de M. le Ministre de la marine et des colonies

PAR L'ADMINISTRATION

DU MUSÉUM IMPÉRIAL D'HISTOIRE NATURELLE.

CINQUIÈME ÉDITION.

PARIS

IMPRIMERIE DE L. MARTINET,

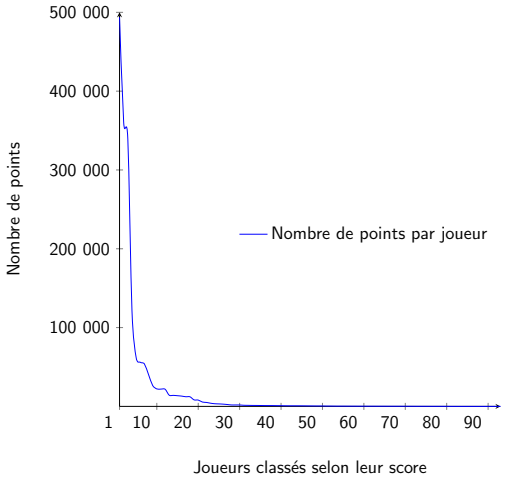
2, RUE MIGNON.

1860

Sciences participatives :

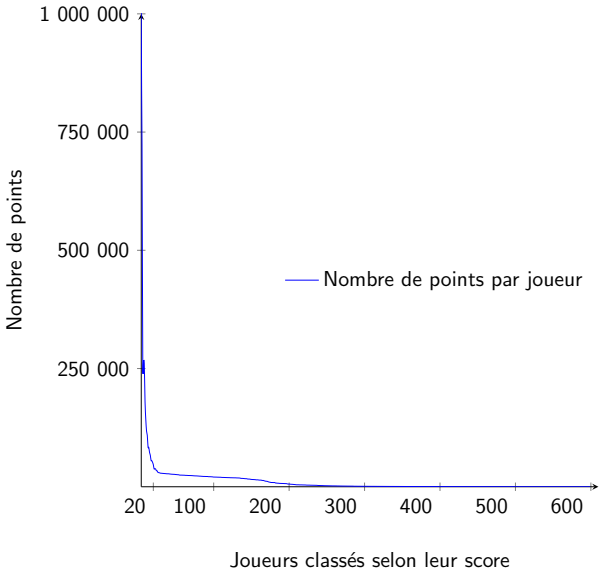
- publiées par le Museum National d'Histoire Naturelle
- première édition en **1824**

Mythe #2 : « Ca implique beaucoup de participants »



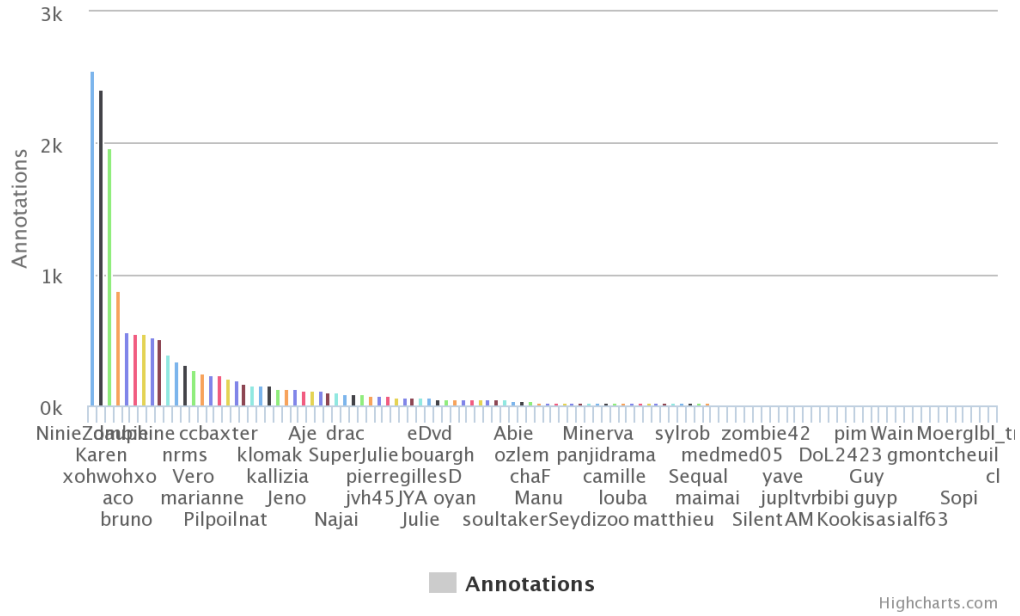
Nombre de joueurs sur *Phrase Detectives* selon le nombre de points gagnés dans le jeu (fév. 2011 - fév. 2012)

Une foule de participants ? JeuxDeMots



Nombre de joueurs sur *JeuxDeMots* selon leur classement dans le jeu (source : http://www.jeuxdemots.org/generateRanking-4.php)

Une foule de participants ? ZombiLingo



Une foule de travailleurs ? [Fort et al., 2011]

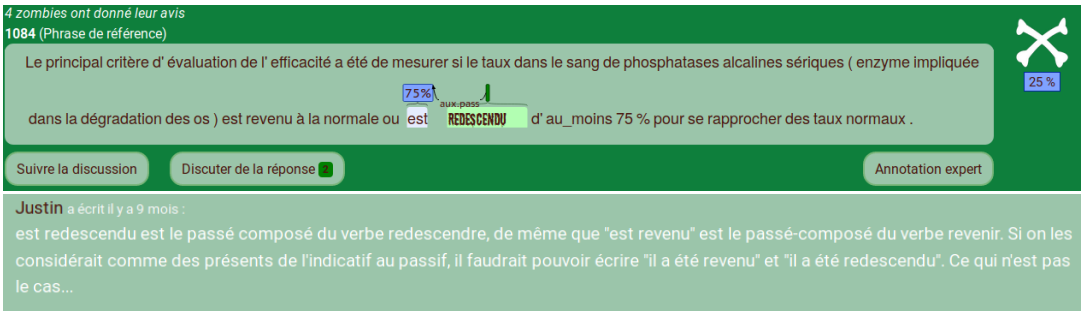
Nombre de *Turkers* actifs sur `Amazon Mechanical Turk` :

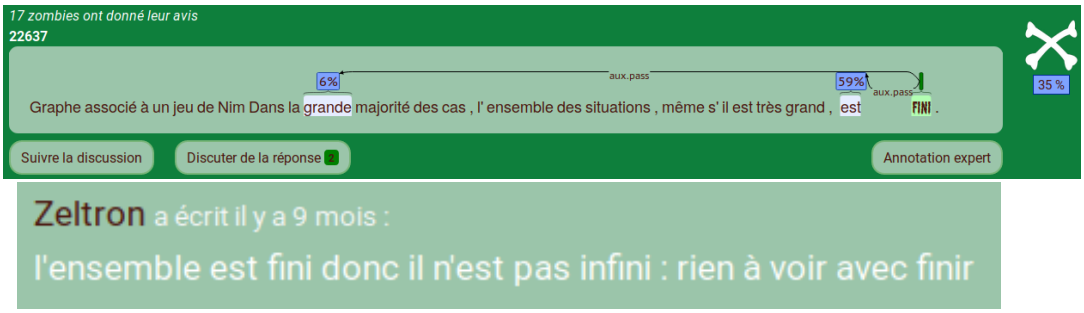
- annoncé : plus de 500 000
- 80 % des tâches (HIT) sont réalisés par les 20 % des *Turkers* les plus actifs [Deneme, 2009]

⇒ travailleurs vraiment actifs (en 2011) : entre 15 059 et 42 912

Mythe #3 : « Ca implique des non-experts »

Extraits du forum de ZombiLingo

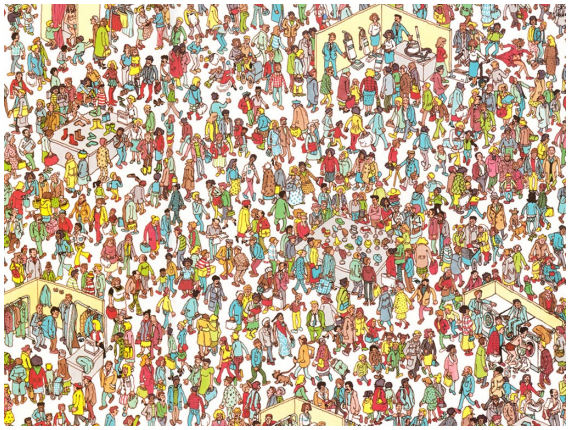




Les productions participatives
Des participants qu'on ne choisit pas
Au-delà des mythes
Redéfinition
Les jeux ayant un but : l'exemple de ZombiLingo
Les incertitudes du modèle
Conclusion et perspectives

Production participative

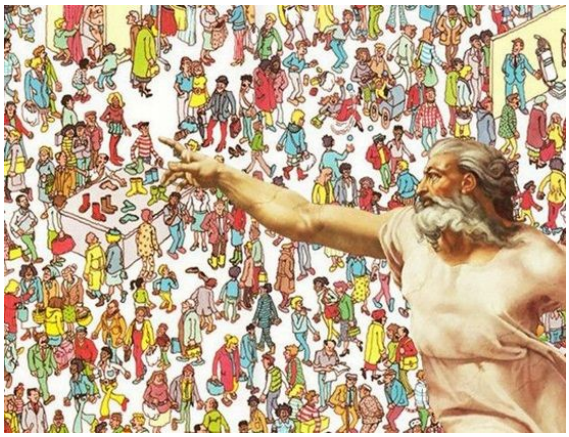
Profiter d'une foule de "non-experts" ?



Production participative

~~Profiter d'une foule de "non-experts" ?~~

→ Trouver/former des experts (de la tâche) dans la foule



Les productions participatives

Les jeux ayant un but : l'exemple de ZombiLingo
- Les jeux ayant un but
- La syntaxe de dépendances, en jeu
- Derrière le rideau
- Résultats

Les incertitudes du modèle

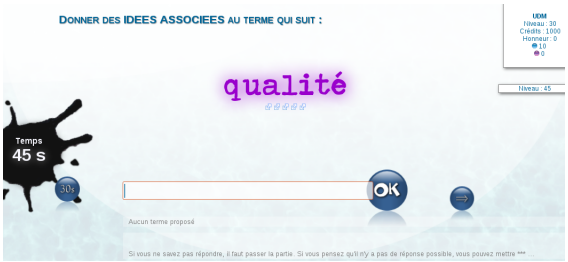
Conclusion et perspectives

JeuxDeMots : jouer à faire des associations d'idées...

... pour créer un réseau lexical [Lafourcade and Joubert, 2008]

Près de 300 million de relations (créées par env. 6 000 joueurs), mises à jour constamment

- jeu par pairs
- des relations de plus en plus complexes, typées
- des challenges
- des procès
- etc

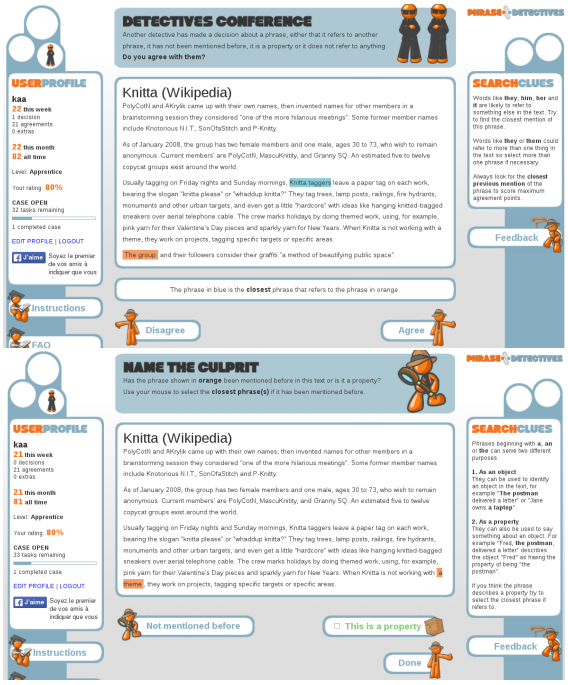


Phrase Detectives : jouer les détectives...

... pour annoter des anaphores [Chamberlain et al., 2008]

3,5 millions de décisions de 45 000 joueurs (2016)

- corpus pré-annoté
- instructions détaillées
- formation
- 2 modes de jeu :
  - annotation
  - validation (correction d'annotations)



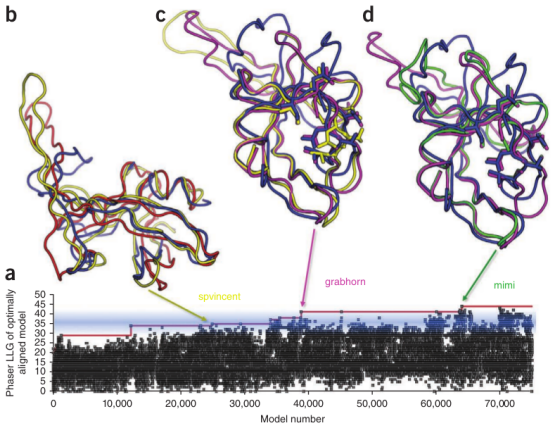
FoldIt : jouer à replier des protéines...

... pour résoudre des problèmes de structure crystalline [Khatib et al., 2011]

Résolution de la structure cristalline de la protéine responsable de la propagation du virus du SIDA chez les macaques rhésus

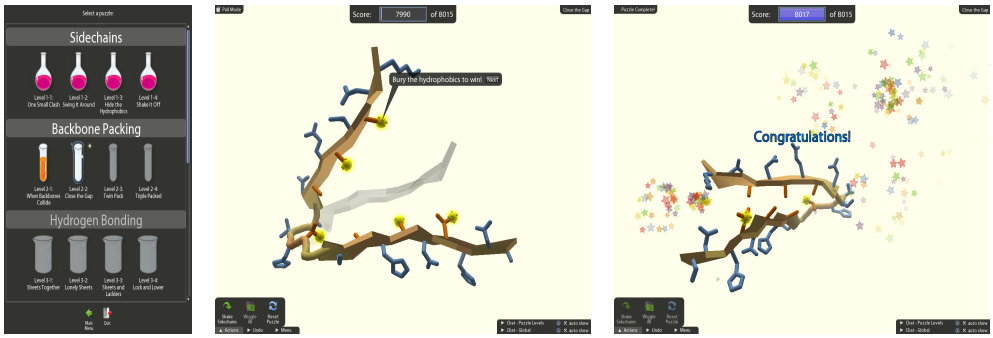
Solution à un problème non résolu depuis plus de 10 ans

- trouvée en quelques semaines
- par une équipe de joueurs
- qui va permettre la création de médicaments antirétroviraux



FoldIt : jouer à replier des protéines. . .

. . . sans aucune connaissance préalable en biochimie [Cooper et al., 2010]



Formation par étapes

- tutoriel décomposé par concepts
- puzzles pour chaque concept
- l'accès aux puzzles suivants n'est octroyé que si le niveau du joueur est suffisant

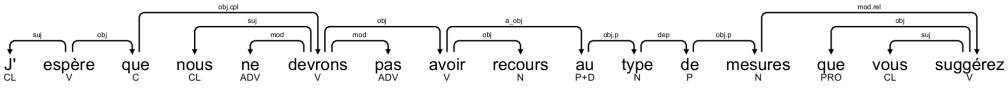
Les productions participatives

Les jeux ayant un but : l'exemple de ZombiLingo
Les jeux ayant un but
La syntaxe de dépendances, en jeu
Derrière le rideau
Résultats

Les incertitudes du modèle

Conclusion et perspectives

Une tâche complexe



- guide d'annotation
  - 29 types de relation
  - approx. 50 pages
- des décisions contre-intuitives (pas de la grammaire d'écoliers, de la linguistique) : `aobj` = *au*

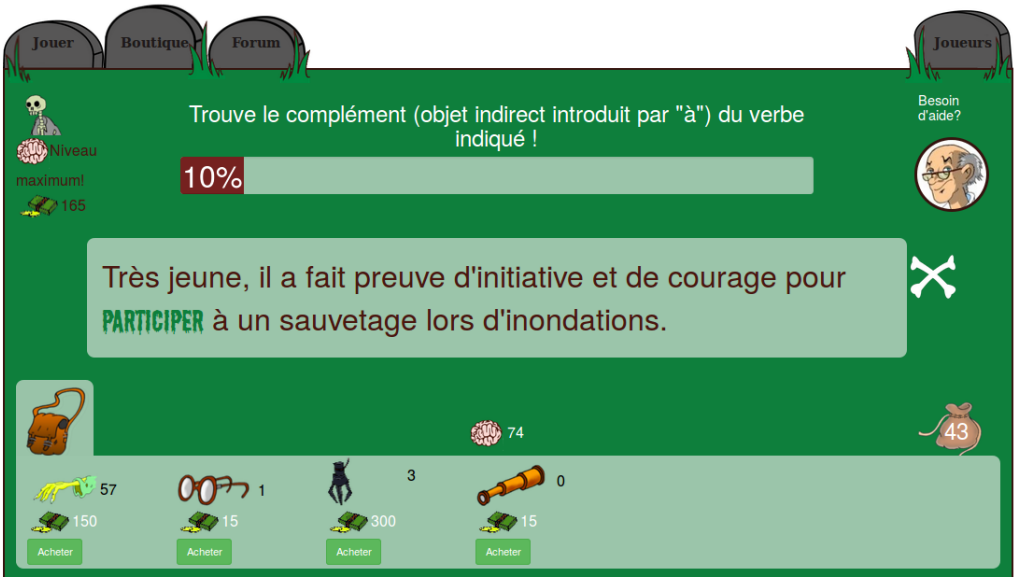
  [...] avoir recours au type de mesures [...]

  c-à-d que la tête de la relation est ici une préposition

→ décomposer la complexité de la tâche [Fort et al., 2012], pas la simplifier !

http://zombilingo.org/

AUXILIAIRE CAUSATIF
AUXILIAIRE DE TEMPS
CHOIX DU CORPUS
POKEMON
NIVEAU 1
SUJET
JEU
FORMATION
Points : 4 980
Annotations : 22/282
NIVEAU 1
AFFIXE
JEU
FORMATION
Points : 0
Annotations : 0/108
NIVEAU 1
DÉTERMINANT
JEU
FORMATION
Points : 0
Annotations : 0/148
NIVEAU 2
COMPLÉMENT EN « DE »
JEU
FORMATION
Points : 0
Annotations : 0/120
NIVEAU 2
COMPLÉMENT EN « À »
JEU
FORMATION
Points : 0
Annotations : 9/157
NIVEAU 2
OBJET DE LA PRÉPOSITION
JEU
FORMATION
Points : 0
Annotations : 0/213
NIVEAU 2
AUXILIAIRE PASSIF
JEU
FORMATION
Points : 598
Annotations : 9/56


Jouer
Boutique
Forum
Joueurs
Niveau maximum!
165
Trouve le complément (objet indirect introduit par "à") du verbe indiqué !
10%
Besoin d'aide?
Très jeune, il a fait preuve d'initiative et de courage pour PARTICIPER à un sauvetage lors d'inondations.
74
43
57
150
Acheter
1
15
Acheter
3
300
Acheter
0
15
Acheter

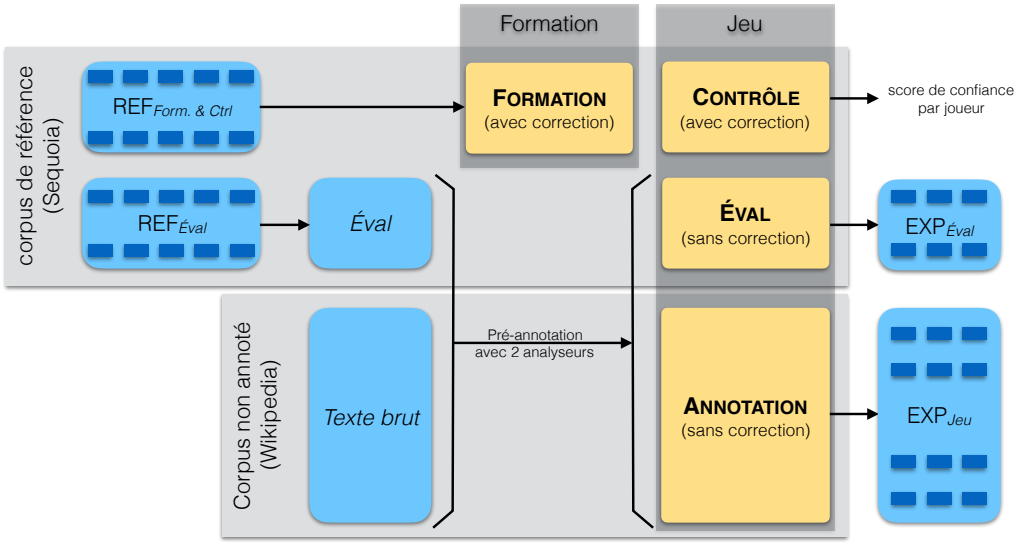
Les productions participatives

Les jeux ayant un but : l'exemple de ZombiLingo
- Les jeux ayant un but
- La syntaxe de dépendances, en jeu
- Derrière le rideau
- Résultats

Les incertitudes du modèle

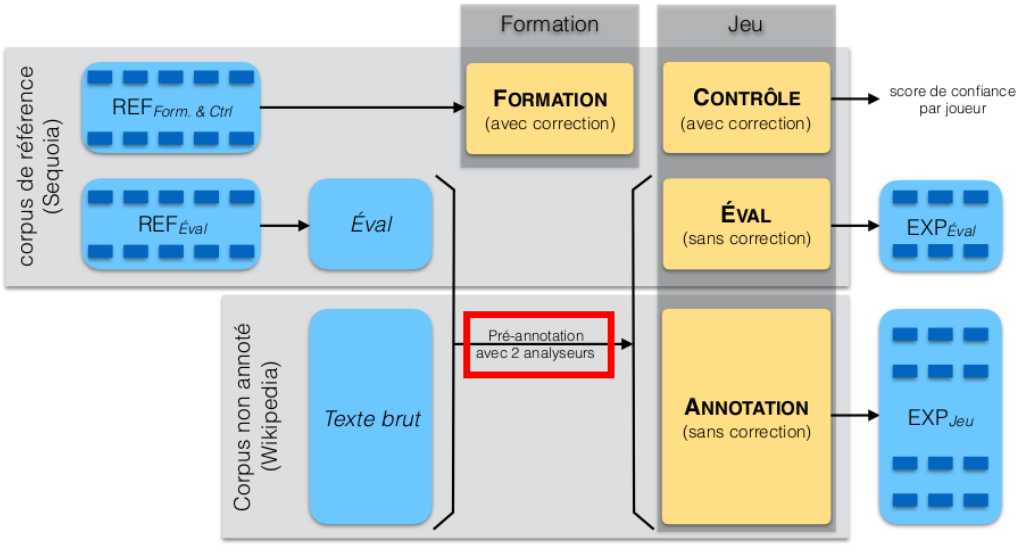
Conclusion et perspectives

Organiser une production de qualité



Prétraitement des données

corpus librement disponibles et distribuables



Prétraitement des données

corpus librement disponibles et distribuables

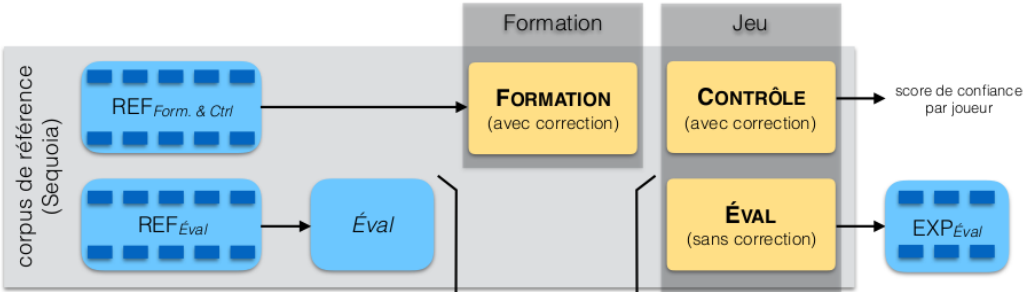
Pré-annotation avec deux parsers

1. un statistique : Talismane [Urieli, 2013]
2. un symbolique, basé sur la ré-écriture de graphes : FrDep-Parse [Guillaume and Perrier, 2015]

→ les joueurs ne jouent que les relations sur lesquelles les deux parsers ne donnent pas les mêmes résultats

Formation, contrôle et évaluation

Référence : 3 099 phrases du corpus Sequoia [Candito and Seddah, 2012]



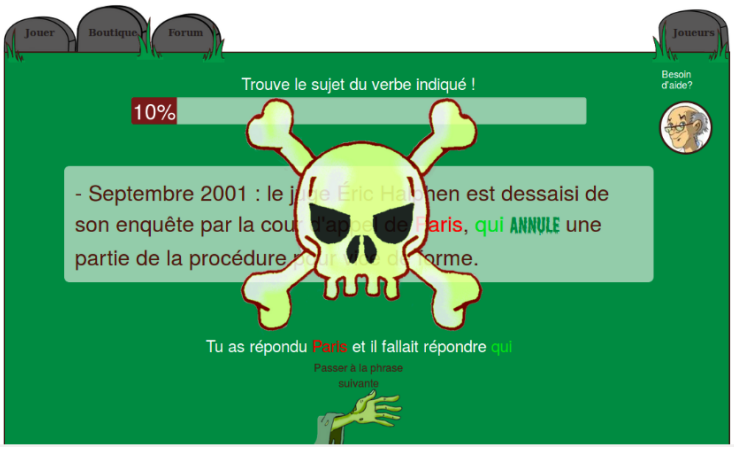
$REF_{Train\&Control}$	REF_{Eval}	non utilisé
50 %	25 %	25 %
1 549 phrases	776 phrases	774 phrases

- $REF_{Train\&Control}$ est utilisé pour former les joueurs
- REF_{Eval} est utilisé comme un corpus brut, pour évaluer les annotations produites

Formation

Obligatoire pour chaque relation

- phrases du corpus $REF_{Train\&Control}$
- retour visuel en cas d'erreur



Gestion de la fatigue cognitive et des joueurs au long court

Mécanisme de contrôle

Des phrases de REF *Train&Control* sont proposées régulièrement

1. si le joueur échoue à trouver la bonne réponse, un retour visuel avec la solution lui est proposé

35 / 59

Gestion de la fatigue cognitive et des joueurs au long court

Mécanisme de contrôle

Des phrases de REF *Train&Control* sont proposées régulièrement

1. si le joueur échoue à trouver la bonne réponse, un retour visuel avec la solution lui est proposé
2. après un certain nombre d'erreurs sur une même relation, le joueur ne peut plus jouer et doit refaire la formation correspondante

36 / 59

Gestion de la fatigue cognitive et des joueurs au long court

Mécanisme de contrôle

Des phrases de REF $_{Train\&Control}$ sont proposées régulièrement

1. si le joueur échoue à trouver la bonne réponse, un retour visuel avec la solution lui est proposé
2. après un certain nombre d'erreurs sur une même relation, le joueur ne peut plus jouer et doit refaire la formation correspondante

→ nous en déduisons un niveau de confiance dans le joueur, pour **cette** relation

Les productions participatives

Les jeux ayant un but : l'exemple de ZombiLingo
- Les jeux ayant un but
- La syntaxe de dépendances, en jeu
- Derrière le rideau
- Résultats

Les incertitudes du modèle

Conclusion et perspectives

Production : taille des corpus créés

Au 10 juillet 2016

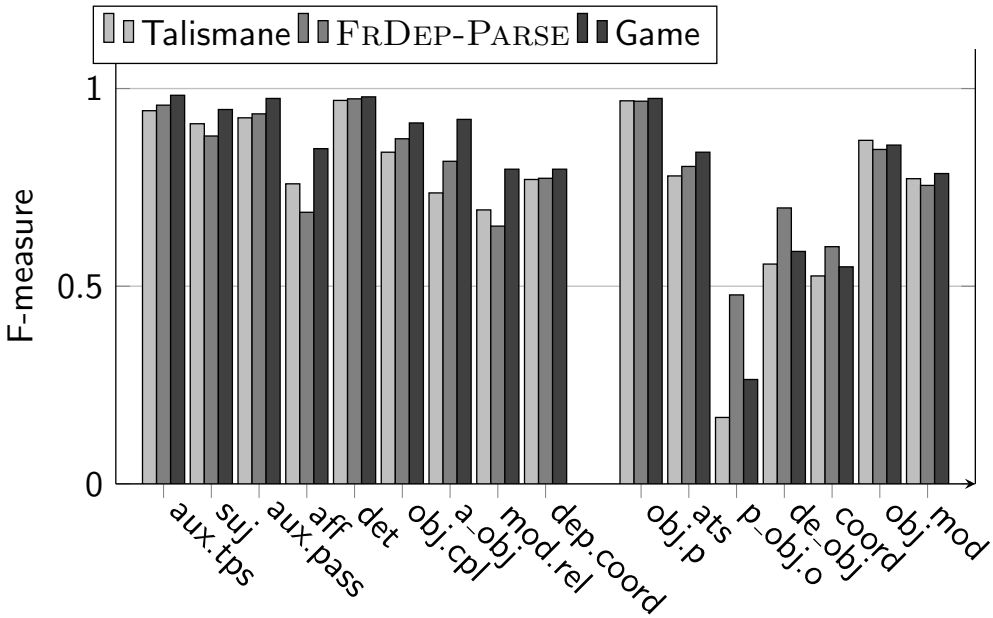
- 647 joueurs (1 460 au 9 octobre 2019)
- ont produit 107 719 annotations (496 462 au 9 octobre 2019)

→ ressource qui évolue constamment !

39 / 59

Évaluation de la qualité

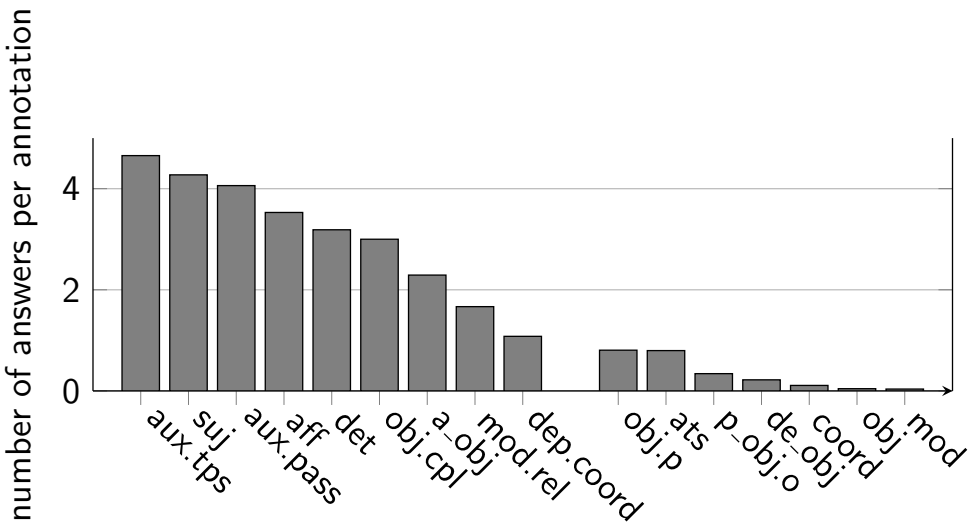
sur le corpus REF_{Eval}



40 / 59

Densité des annotations

sur le corpus REF_{Eval}



→ besoin de **davantage** d'annotations sur certaines relations

Les productions participatives

Les jeux ayant un but : l'exemple de ZombiLingo

Les incertitudes du modèle

- Créer un jeu : un savoir-faire complexe
- Produire des données de qualité
- Motiver la participation : de l'extrinsèque vers l'intrinsèque

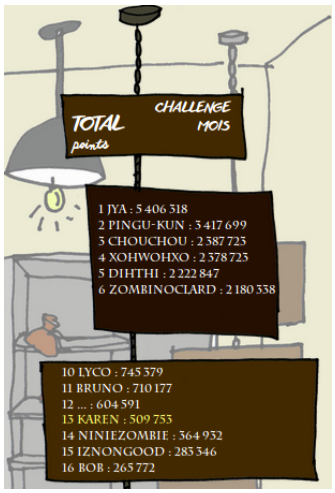
Conclusion et perspectives

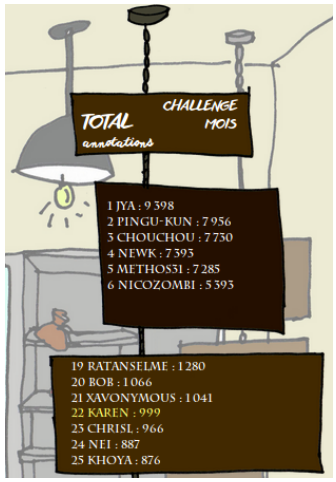
Types de joueurs selon [Bartle, 1996]

Motiver les joueurs en fonction de leur type

- *Achievers* : aiment réussir dans le jeu
- *Explorers* : aiment connaître les coins cachés du jeu
- *Socializers* : aiment interagir avec les autres
- *Killers* : aiment attaquer les autres joueurs

Classements (pour tout le monde)

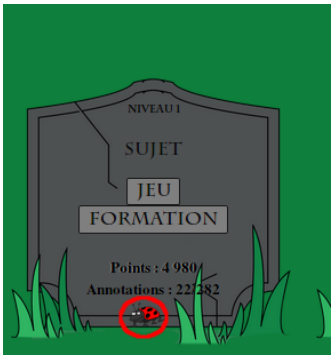




Critères :

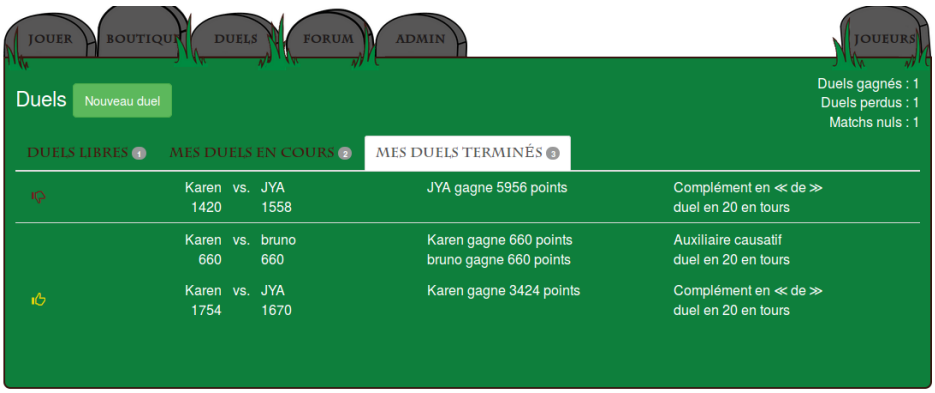
- nombre d'annotations ou de points
- au total, au moins, pendant le défi

Fonctionnalités cachées (pour les explorateurs)



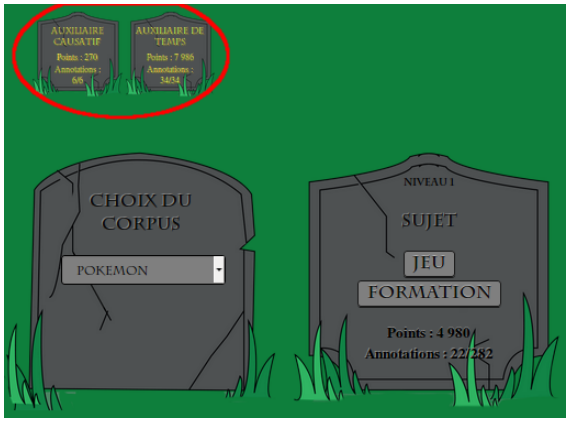
- qui apparaissent aléatoirement
- différents effets : apparition d'objet, lien vers un autre jeu, etc.

Duels (pour les "socialiseurs" (et les tueurs ?))



- sélectionner un ennemi
- lui proposer un duel sur un type de relation

Badges (?) (pour les collectionneurs)



- jouer toutes les phrases pour une relation
- jouer toutes les phrases pour un corpus

Les productions participatives

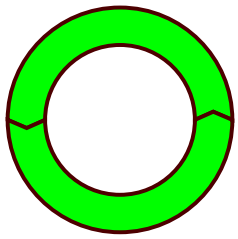
Les jeux ayant un but : l'exemple de ZombiLingo

Les incertitudes du modèle
- Créer un jeu : un savoir-faire complexe
- Produire des données de qualité
- Motiver la participation : de l'extrinsèque vers l'intrinsèque

Conclusion et perspectives

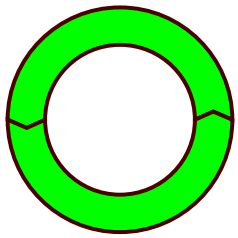
Données de qualité *vs* Fonctionnalités fun

préserver le cercle vertueux n'est pas si facile



Données de qualité *vs* Fonctionnalités fun

préserver le cercle vertueux n'est pas si facile

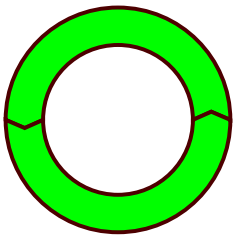


phrase qui disparaît soudainement dans `ZombiLingo` :

+ le joueur est surpris : fun !
- le joueur clique n'importe où : ressource de mauvaise qualité

Données de qualité *vs* Fonctionnalités fun

préserver le cercle vertueux n'est pas si facile



phrase qui disparaît soudainement dans `ZombiLingo` :

- \+ le joueur est surpris : fun !
- \- le joueur clique n'importe où : ressource de mauvaise qualité

joueur qui a trouvé une faille dans `JeuxDeMots` [Lafourcade and Joubert, 2008] pour obtenir du temps :

- \+ crée de la meilleure donnée : bonne qualité
- \- génère de l'envie et de la colère dans la communauté de joueurs : mauvais pour le jeu

Les productions participatives

Les jeux ayant un but : l'exemple de ZombiLingo

Les incertitudes du modèle

Créer un jeu : un savoir-faire complexe

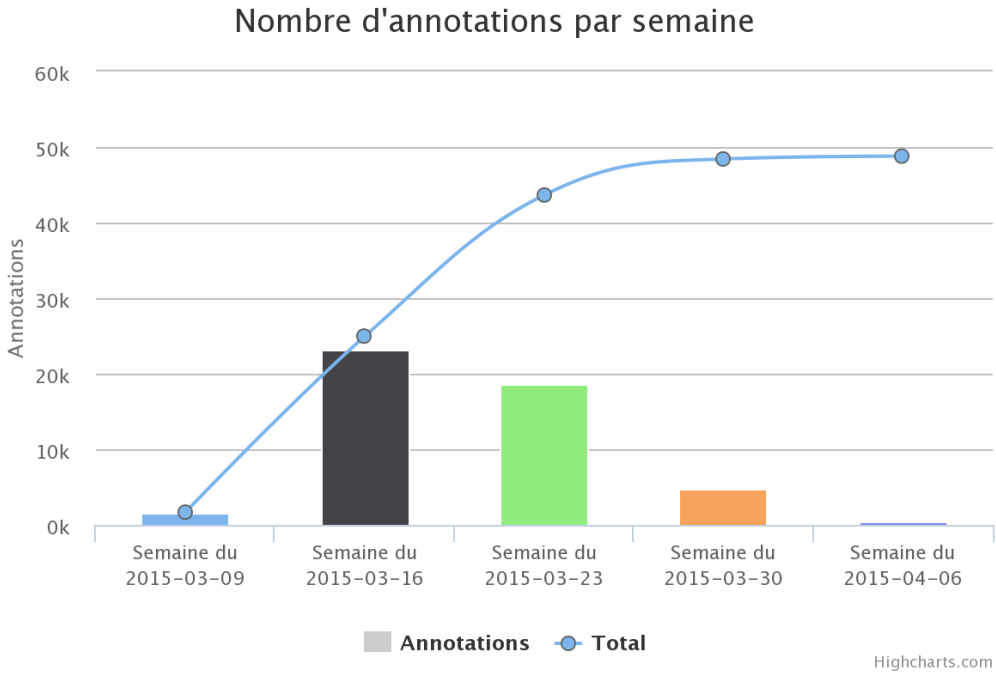
Produire des données de qualité

Motiver la participation : de l'extrinsèque vers l'intrinsèque

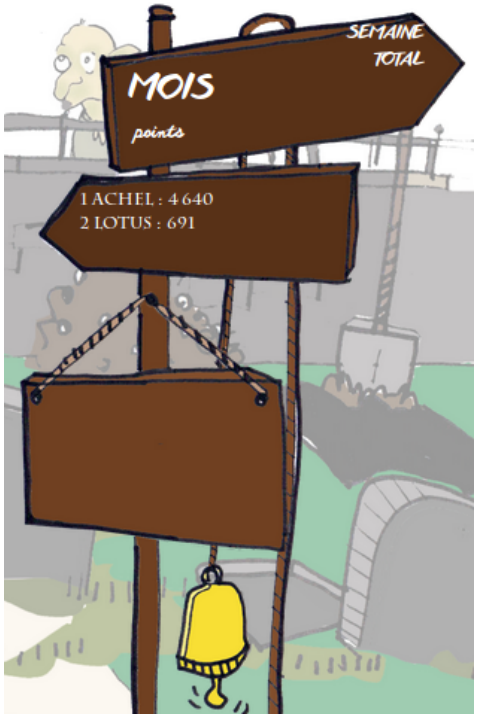
Conclusion et perspectives

Extrinsèque : à relancer constamment

Exemple de ZombiLingo (influence de la semaine de la langue française)



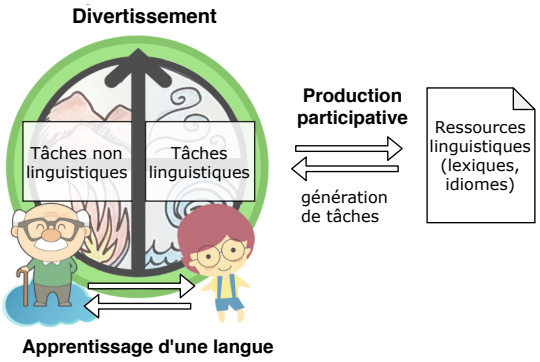
Quand on ne relance pas



Intrinsèque : apprentissage d'une langue

favoriser la transmission inter-générationnelle en produisant des données

Katana et Grand Guru, le jeu des mots perdus



55 / 59

Les productions participatives

Les jeux ayant un but : l'exemple de ZombiLingo

Les incertitudes du modèle

Conclusion et perspectives

56 / 59

Jouir de la foule est un art [Baudelaire, ≪ Les foules ≫]

Faire réaliser des tâches complexes nécessite de :

- connaître les dimensions de complexité de la tâche

→ pour outiller à bon escient

- former les joueurs et les évaluer

→ pour les sélectionner

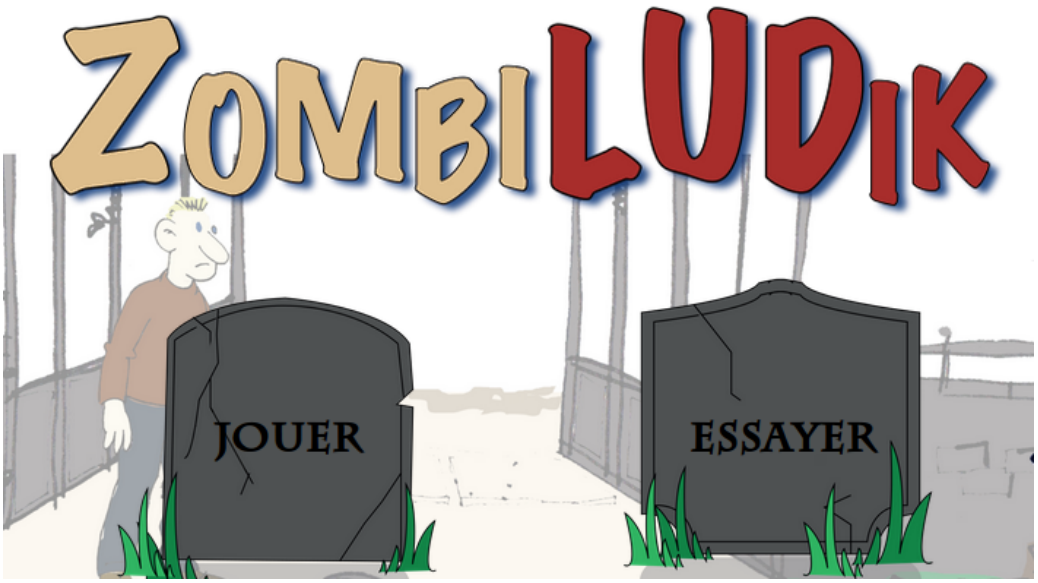
Obtenir des ressources de qualité nécessite de :

- déterminer les moyens et les formes de l'évaluation
- créer un cercle vertueux

57 / 59

Bientôt sur vos écran !

Le jeu pour Universal Dependencies : `http://zombiludik.org/`



58 / 59

https://github.com/zombilingo

http://zombilingo.org/export

59 / 59

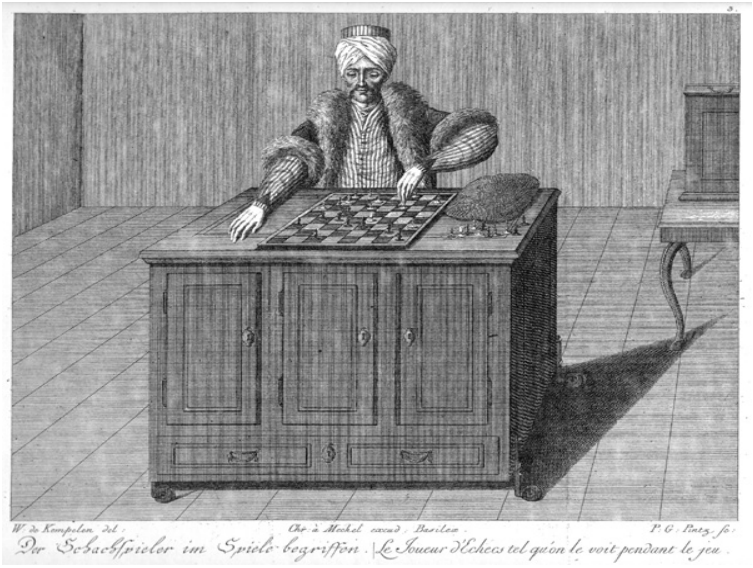
Appendix

Amazon Mechanical Turk : une plate-forme de légendes

Bibliographie

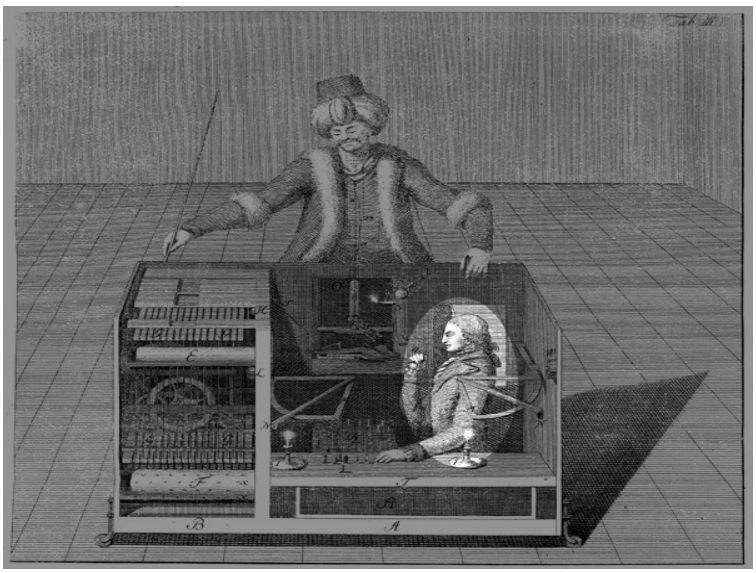
Le «Turc mécanique» de von Kempelen

Un joueur d'échecs mécanique créé par J. W. von Kempelen en 1770 :



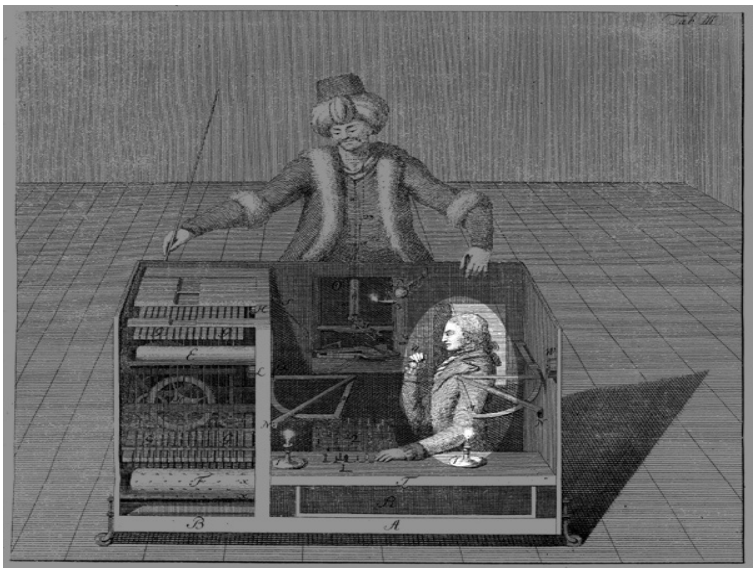
Le «Turc mécanique» de von Kempelen

En fait, un maître d'échecs était caché dans la machine :



Le «Turc mécanique» de von Kempelen

C'est l'intelligence artificielle **artificielle** !



Et Amazon créa AMT

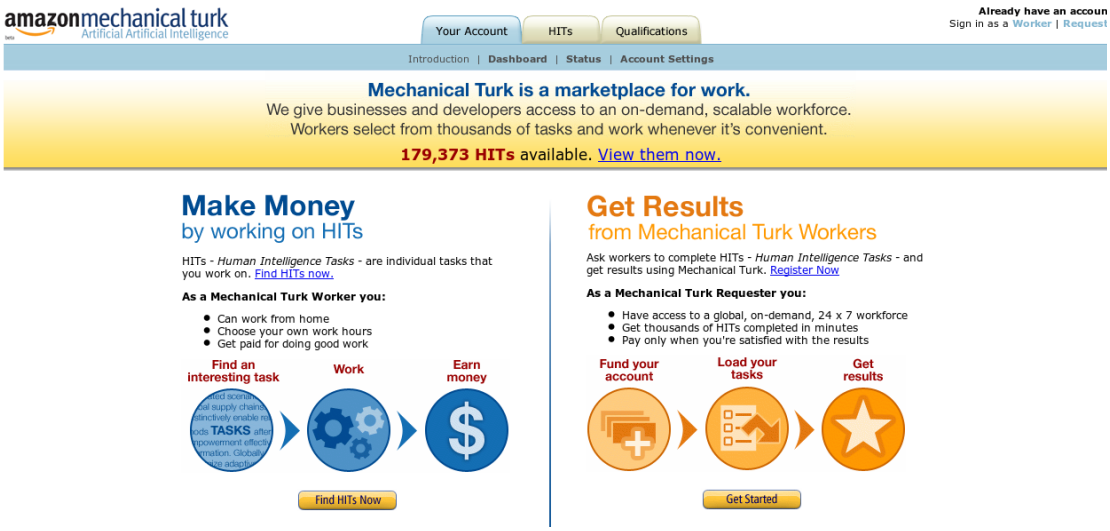
Amazon crée une pour ses propres besoins

plate-forme de travail parcellisé

et en ouvre l'accès en 2005 (moyennant 10 % des transactions)

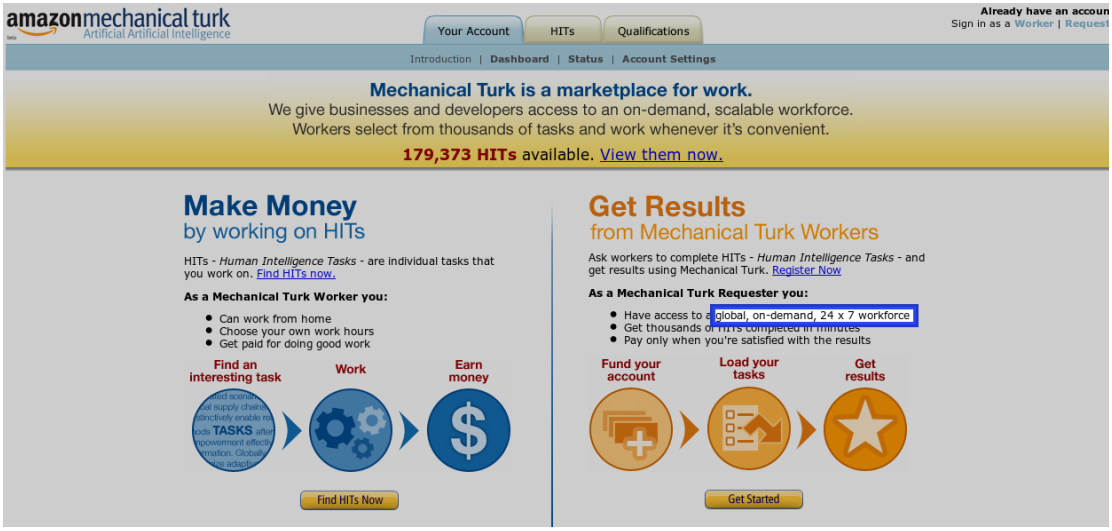
Amazon Mechanical Turk

MTurk



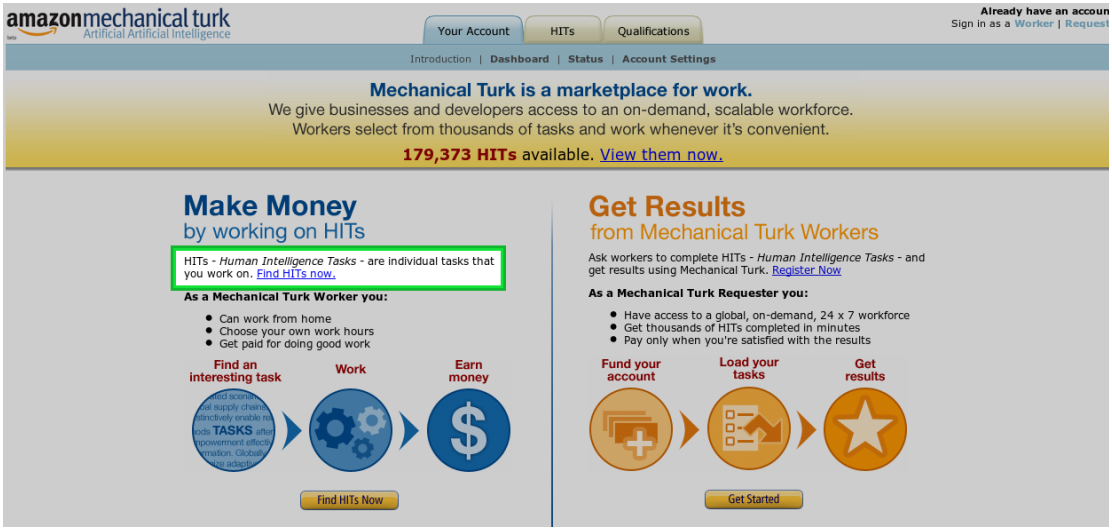
Amazon Mechanical Turk

MTurk est une plate-forme de myriadisation : le travail est *externalisé* via le Web et réalisé par de nombreuses personnes (la *foule*), ici les **Turkers**



Amazon Mechanical Turk

MTurk est une plate-forme de myriadisation du travail parcellisé : les tâches sont découpées en sous-tâches (HIT) et leur exécution est payée par les **Requesters**

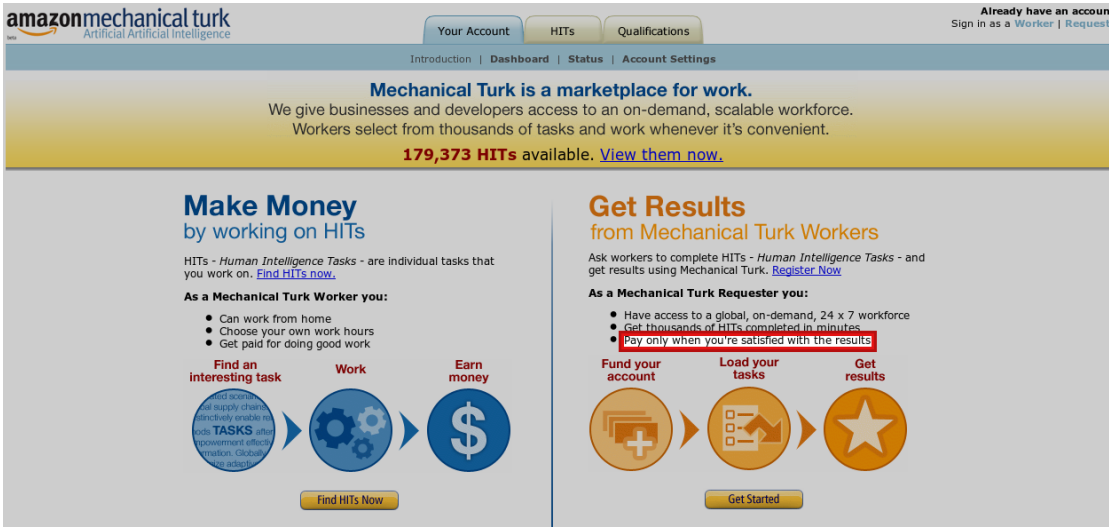


Amazon Mechanical Turk

MTurk est une plate-forme de myriadisation du travail parcellisé : les tâches sont découpées en sous-tâches (HIT) et leur exécution est payée.

Amazon Mechanical Turk

MTurk est une plate-forme de myriadisation du travail parcellisé : les tâches sont découpées en sous-tâches (HIT) et leur exécution est payée.



Caractéristiques d'AMT

Rémunération :

- à la tâche (*illégal* en France sauf (rares) exceptions) : moins de 2 $/h
- pas de relation explicite entre les *Turkers* et les *Requesters*

Tâches :

- nouveaux usages : par exemple, des créations artistiques, comme `http://www.thesheepmarket.com/`
- des tâches traditionnellement réalisées par des employés salariés : transcription, traduction (agences LDC, ELDA), etc

AMT : le rêve devenu réalité ?

Cheap and Fast — But is it Good?
Evaluating Non-Expert Annotations for Natural Language Tasks

Rion Snow[†] **Brendan O'Connor**[‡] **Daniel Jurafsky**[§] **Andrew Y. Ng**[†]

[†]Computer Science Dept.
Stanford University
Stanford, CA 94305
{rion,ang}@cs.stanford.edu

[‡]Dolores Labs, Inc.
832 Capp St.
San Francisco, CA 94110
brendano@doloreslabs.com

[§]Linguistics Dept.
Stanford University
Stanford, CA 94305
jurafsky@stanford.edu

[Snow et al., 2008]

AMT : le rêve devenu réalité ?

Cheap and Fast — But is it Good?
Evaluating Non-Expert Annotations for Natural Language Tasks

Rion Snow[†] **Brendan O'Connor**[‡] **Daniel Jurafsky**[§] **Andrew Y. Ng**[†]

[†]Computer Science Dept.
Stanford University
Stanford, CA 94305
{rion,ang}@cs.stanford.edu

[‡]Dolores Labs, Inc.
832 Capp St.
San Francisco, CA 94110
brendano@doloreslabs.com

[§]Linguistics Dept.
Stanford University
Stanford, CA 94305
jurafsky@stanford.edu

[Snow et al., 2008]

C'est très peu cher, rapide, de bonne qualité
et c'est un hobby pour les *Turkers* !

AMT permet de réduire les coûts

Très basse rémunération ⇒ coûts faibles ? Oui, mais...

- coût de mise au point de l'interface
- coût de création de protections contre les spammers
- coût de validation et de post-traitement

certaines tâches (par exemple, la traduction du pachto vers l'anglais) génèrent des coûts similaires aux coûts habituels dans le domaine, du fait du manque de Turkers qualifiés [Novotney and Callison-Burch, 2010].

AMT permet de produire des ressources de qualité ?

- permet de produire des ressources de qualité dans certains cas précis (par exemple, la transcription simple)
- mais :
  - la qualité est insuffisante lorsque la tâche est complexe (par exemple, le résumé [Gillick and Liu, 2010])
  - l'interface d'AMT pose parfois problème [Tratz and Hovy, 2010]
  - les *Turkers* posent parfois problème (tricheurs, spammers)
  - le modèle de rémunération à la tâche pose problème [Kochhar et al., 2010]
- pour certaines tâches simples les outils de TAL produisent de meilleurs résultats qu'AMT [Wais et al., 2010].

AMT : un passe-temps pour les Turkers ?

[Ross et al., 2010, Ipeirotis, 2010] montre que :

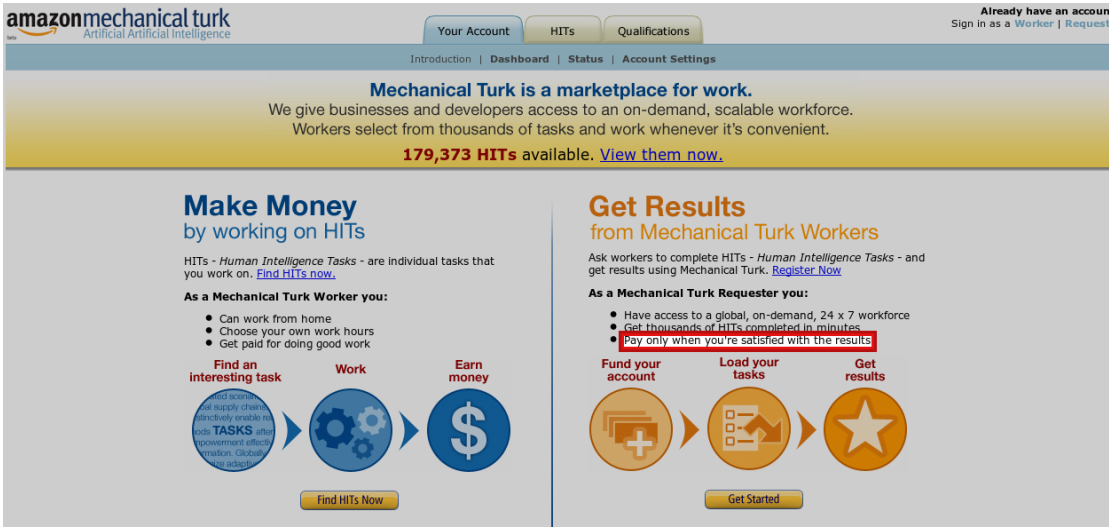
- Turkers sont avant tout motivés par l'**argent** (91 %) :
  - 20 % considèrent AMT comme leur source de revenu primaire ;
  - 50 % comme leur source de revenu secondaire ;
  - l'aspect loisir n'est important que pour une minorité (30 %).
- 20 % des Turkers passent plus de 15 h par semaine sur AMT, et contribuent à 80 % des tâches.
- le salaire horaire moyen observé est inférieur à 2 $.

Est-ce qu'AMT est éthique et/ou légal ?

Éthique :

- pas d'identification : pas de lien officiel entre *Requesters* et *Turkers* et entre *Turkers*
- pas de possibilité de se syndiquer, pour protester contre des manquements des *Requesters* ou ester en justice
- pas de salaire minimum (< 2 \$/h en moyenne)
- possibilité de refuser de payer les *Turkers*

Est-ce qu'AMT est éthique et/ou légal ?



Est-ce qu'AMT est éthique et/ou légal ?

Légalité :

- accord de licence d'`Amazon` : les *Turkers* sont considérés comme des travailleurs indépendants ⇒ ils sont supposés se déclarer comme tels et payer les cotisations afférentes
- illusoire, vus le niveau de rémunération

⇒ les États perdent une source de revenus légitime

Bartle, R. (1996).
Hearts, clubs, diamonds, spades : Players who suit MUDs.
The Journal of Virtual Environments, 1(1).

Candito, M. and Seddah, D. (2012).
Le corpus Sequoia : annotation syntaxique et exploitation pour l'adaptation d'analyseur par pont lexical.
In Proceedings of the Traitement Automatique des Langues Naturelles (TALN), Grenoble, France.

Chamberlain, J., Poesio, M., and Kruschwitz, U. (2008).
Phrase Detectives : a web-based collaborative annotation game.
In Proceedings of the International Conference on Semantic Systems (I-Semantics'08), Graz, Austria.

Cooper, S., Treuille, A., Barbero, J., Leaver-Fay, A., Tuite, K., Khatib, F., Snyder, A. C., Beenen, M., Salesin, D., Baker, D., and Popović, Z. (2010).
The challenge of designing scientific discovery games.

In Proceedings of the Fifth International Conference on the Foundations of Digital Games, FDG '10, pages 40–47, New York, NY, USA. ACM.

Deneme (2009).
How many turkers are there ?
http ://groups.csail.mit.edu/uid/deneme/.

Fort, K., Adda, G., and Cohen, K. B. (2011).
Amazon Mechanical Turk : Gold mine or coal mine ?
Computational Linguistics (editorial), 37(2) :413–420.

Fort, K., Nazarenko, A., and Rosset, S. (2012).
Modeling the complexity of manual annotation tasks : a grid of analysis.
In International Conference on Computational Linguistics (COLING), pages 895–910, Mumbai, India.

Gillick, D. and Liu, Y. (2010).
Non-expert evaluation of summarization systems is risky.

In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk, CSLDAMT '10, pages 148–151, Stroudsburg, PA, USA. Association for Computational Linguistics.

Guillaume, B. and Perrier, G. (2015).
Dependency Parsing with Graph Rewriting.
In
Proceedings of IWPT 2015, 14th International Conference on Parsing pages 30–39, Bilbao, Spain.

Howe, J. (2006).
The rise of crowdsourcing.
Wired Magazine, 14(6).

Ipeirotis, P. (2010).
The new demographics of mechanical turk.
http ://behind-the-enemy-lines.blogspot.com/2010/03/new-demographics-of-mechanical-turk.html.

Khatib, F., DiMaio, F., Cooper, S., Kazmierczyk, M., Gilski, M., Krzywda, S., Zabranska, H., Pichova, I., Thompson, J., Popović, Z., et al. (2011).
Crystal structure of a monomeric retroviral protease solved by protein folding game players.
Nature structural & molecular biology, 18(10) :1175–1177.

Kochhar, S., Mazzocchi, S., and Paritosh, P. (2010).
The anatomy of a large-scale human computation engine.
In Proceedings of Human Computation Workshop at the 16th ACM SIKDD Conference on Knowledge Discovery and Data Mining, KDD 2010, Washington D.C.

Lafourcade, M. and Joubert, A. (2008).
JeuxDeMots : un prototype ludique pour l'émergence de relations entre termes.
In Proceedings of the Journées internationales d'Analyse statistique des Données Textuelles (JADT), Lyon, France.

Novotney, S. and Callison-Burch, C. (2010).

Cheap, fast and good enough : automatic speech recognition with non-expert transcription.
In Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), HLT'10, pages 207–215, Stroudsburg, PA, USA. Association for Computational Linguistics.

Ross, J., Irani, L., Silberman, M. S., Zaldivar, A., and Tomlinson, B. (2010).
Who are the crowdworkers ? : shifting demographics in mechanical turk.
In Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems, CHI EA '10, pages 2863–2872, New York, NY, USA. ACM.

Snow, R., O'Connor, B., Jurafsky, D., and Ng., A. Y. (2008).
Cheap and fast - but is it good ? evaluating non-expert annotations for natural language tasks.
In Proceedings of EMNLP 2008, pages 254–263.

Tratz, S. and Hovy, E. (2010).

A taxonomy, dataset, and classifier for automatic noun compound interpretation.
In Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics, pages 678–687, Uppsala, Sweden. Association for Computational Linguistics.

Urieli, A. (2013).
Robust French syntax analysis : reconciling statistical methods and linguistic knowledge in the Talismane toolkit.
PhD thesis, Université de Toulouse II le Mirail, France.

Wais, P., Lingamneni, S., Cook, D., Fennell, J., Goldenberg, B., Lubarov, D., Marin, D., and Simons, H. (2010).
Towards building a high-qualityworkforce with mechanical turk.
In Proceedings of Computational Social Science and the Wisdom of Crowds (NIPS).

1.2 Prediction of Missing Semantic Relations in Lexical-Semantic Network using Random Forest Classifier

Kevin Cousot, Mehdi Mirzapour, Waleed Ragheb

Prediction of Missing Semantic Relations in Lexical-Semantic Network using Random Forest Classifier

Kevin Cousot[1], Mehdi Mirzapour[1], Waleed Ragheb[1]

[1] Université de Montpellier, LIRMM, CNRS, Montpellier, France
{kevin.cousot, mehdi.mirzapour, waleed.ragheb}@lirmm.fr

Résumé. Cette étude porte sur la prédiction de six relations sémantiques manquantes (telles que is_a et has_part) entre deux nœuds de RezoJDM, un réseau lexico-sémantique pour le français. Le résultat de cette prédiction est un ensemble de paires dans lesquelles les premières entrées sont des relations sémantiques et les secondes sont les probabilités d'existence de la relation. En raison de l'énoncé du problème, nous avons choisi d'utiliser un classifieur de forêt aléatoire pour le résoudre. Nous exploitons RezoJDM pour en extraire les relations sémantiques et construire nos données d'entraînement/test. Nous expliquons en quoi les idées développées peuvent être utilisées après l'utilisation de l'approche node2vec dans la phase de feature extraction. Nous montrons finalement comment cette approche conduit à des résultats prometteurs.

Mots-clés. Apprentissage automatique, apprentissage supervisé, prédiction de relations sémantiques, classifieur de forêts aléatoires, réseau lexico-sémantique, traitement du langage naturel.

Abstract. This study focuses on the prediction of missing six semantic relations (such as is_a and has_part) between two given nodes in RezoJDM a French lexical-semantic network. The output of this prediction is a set of pairs in which the first entries are semantic relations and the second entries are the probabilities of existence of such relations. Due to the statement of the problem we choose the random forest (RF) predictor classifier approach to tackle this problem. We take for granted the existing semantic relations, for training/test dataset, gathered and validated by crowdsourcing. We describe how all of the mentioned ideas can be followed after using node2vec approach in the feature extraction phase. We show how this approach can lead to acceptable results.

Keywords. Machine Learning, Supervised Learning, Semantic Relation Prediction, Random Forest Classifier, Lexical-Semantic Network, Natural Language Processing, Computational Semantics.

1 Introduction

Prediction of semantic relations in natural language texts is an important task in Natural Language Processing and Computational Linguistics/ Semantics. Scaling up semantic analysis on natural language level needs

some sort of semantic relations prediction at some point [8]. Obviously, it is not practical to define all the possible relations for all the possible words or nodes in a network. This naturally leads language models to some sort of trade-off for balancing manually definitions of semantic relations in one hand and prediction of missing relations between unconnected words/nodes in the other hand.

This study specifically focuses on missing semantic relations for French language. Current state-of-the-art methodologies [7], [8] suggest different approaches such as proximity of nodes, similarity-based techniques, maximum likelihood and probabilistic models for solving this kind of the problem. In general, our approach relies on supervised machine learning and tries to introduce the best combination of existing graph embeddings technique [3] to solve the problem of semantic relation prediction for French language.

We have used RezoJDM [5] a French lexical-semantic network which have helped us significantly to create our training/test datasets. To best of our knowledge, no study has been focused on RezoJDM for semantic relation predictions on French vocabularies. The motivation behind this paper is grounding a feasibility study in feature extraction and semantic link predictions on the level of RezoJDM the biggest and state-of-the-art French lexical-semantic network using data-oriented approaches.

Famous supervised machine learning approaches match quite well for the mentioned trade-off that can be re-interpreted as the size of training/test data and the unseen data. To build our training/test dataset, two general approaches can be considered: (i) to manually label the existing corpora with semantic relations (ii) to use state-of-the-art French lexical-semantic networks such as RezoJDM (section 2) which are already built up by crowdsourcing and automatic validations. We focus only on semantic relations; precisely speaking, in this phase of research, we have restricted our focus on the six most important relations, namely: r_raff_sem(relation which links a polysemous term to its different senses), r_syn(relation of synonymy), r_isa (relation of hypernymy, i.e., linking a generic term to a specific instance of it), r_hypo (relation of hyponymy), r_lieu(location relation, i.e., the place where an action / an object can take place / is usually found), r_agent(agent relation) and r_agent-1(inverse of r_agent).

The inputs of our model are set of pairs representing two nodes/words with missing semantic relations in JDM and the output of this prediction is a set of pairs in which the first entries are semantic relations and the second entries are the probabilities of existence of such relations for each pair in the input set. Due to the statement of the problem we choose the random forest (RF) predictor classifier approach [1], [2] to tackle this task. We take for granted the existing semantic relations in RezoJDM, for training/test dataset, gathered and validated by crowdsourcing.

The rest of the paper is organized as follow: Section 2 is a short introduction on RezoJDM. Section 3 is about three-fold training/test datasets building methodology and its feature extraction. Section 4 explains our RF modeling, its parameters setting and the accuracy of the model. In the last section, we conclude our paper and discuss possible future works.

It is worth mentioning that the datasets, saved graphed embedding and python implementation of our RF model is accessible here: https://github.com/mehdi-mirzapour/JDM_Graph_Embedding

2 RezoJDM : A French Lexical-Semantic Network

RezoJDM (http://www.jeuxdemots.org) is a lexical-semantic network that is built by means of online games, so-called Games With A Purpose (GWAP), launched from 2007 [5]. It is a large graph-based network, in constant evolution, containing more than 310,000 terms connected by more than 6.5 million relations. JDM is a graph structure composed of nodes (vertices) and links. Each node in JDM is a 3-tuple: <name, type, weight> and the link is a 4-tuple: <start-node, type, end-node, weight>.

The name is a string denoting the term. The node type is an encoding referring to the information of the node. For instance, a node can be a term or a Part of Speech. The link type refers to the relation considered. A node weight is interpreted as a value referring to the frequency of usage of the term. The relation weight, similarly, refers to the strength of the relation. The link types are divided into predetermined list of the categories, namely: lexical, ontological, associative and predicative relations.

Validation of the relations in RezoJDM which are anonymously given by a player is made also anonymously by other players. A relation is considered valid if it is given by at least one pair of players. A game

happens between two players. A first player-- let's say player A-- begins a game with prompting an instruction with a term T randomly picked from the database. The player A has then a limited time to answer which answer is applicable to the term T. The number of propositions which he can make is limited to letting the player not just type anything as fast as possible, but to have to choose the proper one. The same term with the same instruction is later proposed to another player B; the process is then identical. To increase the playful aspect, for any common answer in the propositions of both players, they receive a given number of points. At the end of a game, propositions made by the two players are shown, as well as the intersection between these terms and the number of points they win.

3 Building Training/Test Datasets from RezoJDM Graph

In order to build training/test datasets from RezoJDM, we have taken following three stages:

3.1 Subgraphs Selection from RezoJDM

Since JDM is a very big network, we have selected only a part of the existing subgraphs of the network for our training/test dataset in this study. The nodes are restricted just to 2 types of "n_term" and "n_form". The edges are restricted to types of "r_raff_sem", "r_syn|", "r_isa", "r_hypo", "r_lieu", "r_agent-1", "r_agent". Experimental trials suggest the threshold of 60 for the weight attributes of all the mentioned relations.

3.2 Feature Learning from RezoJDM

For feature learning from RezoJDM network we have used node2vec approach [3] which is a mapping of nodes to a low-dimensional space of features that maximizes the likelihood of preserving network neighborhoods of nodes. Node2vec defines a flexible notion of a node's network neighborhood and design a biased random walk procedure, which efficiently explores diverse neighborhoods.

Experiments demonstrate that node2vec outperforms state-of-the-art methods by up to 26.7% on multi-label classification and up to 12.6% on link prediction [3]. In this paper, we have avoided to go through

technicalities in the node2vec approach. The parameter settings used for node2vec are:

Number of dimensions=128
Length of walk per source= 80
Number of walks per source=1
Return hyperparameter=1
In-out hyperparameter=1
Vector dimension size=20

After applying the node2vec algorithm, we ended up creating vectors with size parameter of 20 for each source and end node in the selected subgraphs. Consequently, 40 columns were built that can properly serve as the input for the further supervised learning algorithm. A cursory investigation on the outcome of our parameter settings shows interesting results: similarity between femme(=woman), homme(=man) is 0.9395609, also the most similar vector to 'femme'(=woman)-'roi'(=king)-'homme'(=man) is 'princesse' (0.99). We will see more in section 4 how our parameter setting can lead to interesting results.

3.3 Creating the Training/Test Datasets

We can build a training/test dataset with source and destination words and the relevant ids plus 40 features learned from node2vec approach (section 3.2). There are six columns related to the labels which include binary semantic relations (section 3.1). The labels show the existence/nonexistence of the relations in the JDM network. We have selected only the edges with weights greater than 80 in final stage. We have end up with around 37,000 rows of data for this phase of our study. It is worth mentioning that the cleaned dataset can be accessed here: (https://github.com/mehdi-mirzapour/JDM_Graph_Embedding/blob/master/cleaned_JDM_dataset.xlsx)

4 Prediction of Missing Semantic Relations by means of Random Forest Classifier

Random Forest (RF) algorithm [1], [2] has been considered as one of the most precise prediction methods for classification and regression. In this section we apply this approach for predicting the semantic relations. We finish with our parameter settings and its performance analysis.

4.1 Parameter Settings for Random Forest Classifier

Random forest (RF) classifier algorithm [1], [2] is an ensemble-learning algorithm that combines the predictions of multiple models to create a more accurate final prediction. There are a number of parameters that needs to be set for our model that we can tweak them. Some of them are related to the individual trees in the model, and change how they are constructed such as *min_samples_leaf, min_samples_split and max_depth, max_leaf_nodes*. We have set all of these parameters to their default values. There are also parameters specific to the random forest that alter its overall construction such as *n_estimators* and bootstrap (or bagging) which is turned on. For our RF Classifier, the *n_estimators* parameter value is optimized using cross-validation to find the values that could best predict the semantic relation status. For each stage, *n_estimators* from 10 to 170 with intervals of length 20 were tested. The following figure shows that the accuracy of model will not increase a lot after setting *n_estimators* to 170. In real practice, we have set this parameter to 70 for all six type of relations to reduce computing.

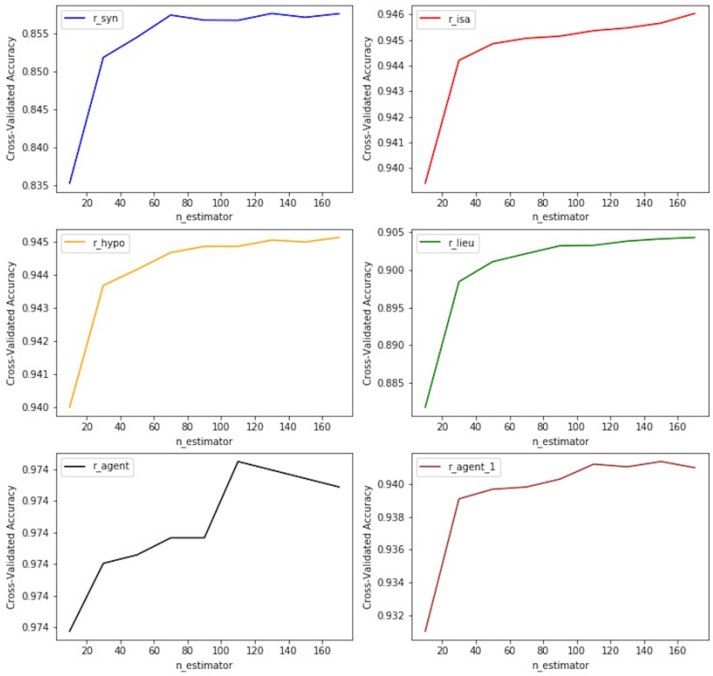


Fig. 1. RF n_estimator parameter setting. The analysis shows that setting the parameter to 100 can properly make the trade-off between performance and accuracy.

4.2 Performance Analysis

The RF algorithm have several advantages comparing to other frameworks [1]. To count some of them: (i) it runs efficiently on large datasets, (ii) it is less sensitive to noise or over-fitting, (iii) it has fewer parameters, and finally (iv) it has better performance. In this subsection we explain some of the metrics related to our RF performance. As it is illustrated in the table (1), recall, precision and F1 scores have good level of qualities for r_syn, r_is, r_lieu and r_agent_1. It is worth mentioning that our reference for following results is based on JDM test-set itself. This might not be considered as a highly reliable approach and can be interpreted as some sort of internal bias. Nevertheless, this can be counted into account in other studies that deals with reliability problem and it is not explored in this study.

Table 1. Recall, Precision, Accuracy and F1 scores after applying the tuned Random Forest for six semantic relations.

	r_syn	r_isa	r_hypo	r_lieu	r_agent_1	r_agent
Recall	0.79	0.54	0.37	0.84	0.61	0.20
Precision	0.84	0.90	0.82	0.87	0.84	0.77
F1 Score	0.81	0.66	0.51	0.85	0.70	0.31

Two remaining relations, namely, r_hypo and r_agent do not show proper recall indicators. This has affected the F1 score, since F1 relies on recall as one of its input factors. Low recall and high precision in our modeling have happened due to the unbalanced classes in our datasets. Tweaking *class_weigh* parameter does not improve much our model. Making balanced datasets needs more computing power and better strategies in our dataset creating phase.

5 Conclusion and Possible Future Works

In this paper, we reported the feasibility study of prediction of missing semantic relations between two given nodes in RezoJDM French lexical-semantic network. We have also provided a methodology for our feature

extraction that is very useful for different natural language processing tasks—regardless of its application for our specific task. We implemented all of the ideas in this paper and we have gained rather satisfactory results: recall 55% and precision 84%, F1 score 64% (percentages are on average for all six types of semantic relations).

We reported an important issue in section 4.2 regarding lacking the balanced datasets. Further enhancement for overcoming this drawback has left for our future studies and needs more consideration on new strategies and balancing the computing resources.

We are planning to extend our study to include the full RezoJDM network with more relations. Moreover, we can apply deeper graph embedding methods to produce unsupervised node features from big graphs and also make use of deeper models rather than the classical machine learning for classification step. Reformulating the problem from binary classification to multilabel and multiclass classification can be done as another extension to our current work.

The introduced workflow in this paper, is intentionally designed for some future projects in computational semantics field. The partial success in this study can pave the way for following future applications. To count some of them, we can mention the following NLP tasks:

(i) In some sentence completion task algorithm [9], we need to eliminate different candidates for completion of a given sentence. One way of doing that is eliminating the candidates that have very low probabilities between the missing words and the existing words in the sentence.

(ii) Finding linguistic coercions in computational linguistics field [6], [10] is another important task for deep semantic analysis and semantic preference modeling.

(iii) Measuring linguistic complexity on syntactic level [11] needs to be enriched with some extra semantic relations on the word level. This task can be definitely enhanced by using semantic relations prediction as we have described in this research.

Acknowledgments. We would like to appreciate Mathieu Lafourcade, Michael Carl and Christian Retoré for their kind supports and discussions.

References

1. Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and Regression Trees. The Wadsworth statistics probability series. Wadsworth International Group, Belmont, CA.

2. Breiman, L.: Random Forests Machine Learning (2001) 45: 5. October 2001, Volume 45, Issue 1, pp 5–32, https://doi.org/10.1023/A:1010933404324

3. Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 855-864). ACM.

4. Ho, Tin Kam.: Random Decision Forests (1995). Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 14–16. pp. 278–282.

5. Lafourcade, M. (2007). Making people play for Lexical Acquisition with the JeuxDeMots prototype. In: SNLP'07: 7th International Symposium on Natural Language Processing. Pattaya, Chonburi, Thailand,p. 7. URL: https://hal-lirmm.ccsd.cnrs.fr/lirmm-00200883.

6. Lafourcade, M., Mery B., Mirzapour M., Moot R., Retoré C. (2018). Collecting Weighted Coercions from Crowd-Sourced Lexical Data for Compositional Semantic Analysis. In: Arai S., Kojima K., Mineshima K., Bekki D., Satoh K., Ohta Y. (eds) New Frontiers in Artificial Intelligence. JSAI-isAI 2017. Lecture Notes in Computer Science, vol 10838. Springer, Cham.

7. Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. Journal of the American society for information science and technology, 58(7), 1019-1031.

8. Lü, L., & Zhou, T. (2011). Link prediction in complex networks: A survey. Physica A: statistical mechanics and its applications, 390(6), 1150-1170.

9. Mirzapour, M. (2017), Finding Missing Categories in Incomplete Utterances. In 24e Conférence sur le Traitement Automatique des Langues Naturelles (TAL), p. 149, Orléans, France.

10. Mirzapour, M. (2018). Modeling Preferences for Ambiguous Utterance Interpretations. Université Montpellier. English. ⟨NNT : 2018MONTS094⟩. ⟨tel-01908642v2⟩

11. Mirzapour M., Prost J-P., Retoré C. (2019). Measuring Linguistic Complexity: Introducing a New Categorial Metric. Logic and Algorithms in Computational Linguistics, Series: Advances in Intelligent Systems and Computing, In press, Studies in Computational Intelligence. ⟨hal-02146506⟩

1.3 La logique hyperdocumentaire dans la structuration de l'information et sa visualisation

Arthur Perret

La logique hyperdocumentaire dans la structuration de l'information et sa visualisation

Arthur Perret[1]

[1] Université Bordeaux Montaigne, MICA EA 4426 10
Esplanade des Antilles, 33607 Pessac CEDEX, France
arthur.perret@u-bordeaux-montaigne.fr

Résumé. La logique hyperdocumentaire correspond à la conception de la documentation sur le modèle du réseau dans les travaux pionniers de Paul Otlet. Elle repose sur une nouvelle théorie du document et de la classification qui tend vers l'encyclopédie. Parce qu'elle est indissociable du graphe, l'hyperdocumentation s'inscrit à la fois dans le champ de l'épistémologie visuelle et des humanités numériques. Le graphe constitue un outil d'exploration réflexif de la logique hyperdocumentaire. Cela nous permet de développer une compréhension pratique du lien entre structuration et visualisation.

Mots-clés. documentation, épistémologie, réflexivité, graphe, corpus

Abstract. Hyperdocumentary logic describes the way documentation was redefined as a network structure in the pioneering works of Paul Otlet. This logic is based on a new theory of documents and classification which leads towards encyclopedia. Because it is inseparable from graphs, hyperdocumentation beckons an approach rooted in visual epistemology and digital humanities. Graphs can be used as tools for the reflexive exploration of hyperdocumentary logic. This allows us to develop a practical comprehension of the link between structuration and visualization.

Keywords. documentation, epistemology, reflexivity, graphs, corpus

1 Introduction

Paul Otlet (1868-1944) est un avocat, bibliographe et militant pacifiste belge considéré comme le fondateur de la documentation moderne. Ses travaux offrent une réponse théorique et pratique au défi informationnel de son époque et de la nôtre. En effet, faisant face à un accroissement considérable de la production écrite (notamment scientifique et technique), il fait le constat de la difficulté immense que représente l'acquisition d'une connaissance globale du monde, non seulement pour un individu

mais également à l'échelle de la société – problématique toujours d'actualité à l'ère d'Internet et des *big data*. Otlet estime qu'il y a un lien direct entre l'ignorance, l'incommunication et la guerre ; son projet est celui de la connaissance universelle pour favoriser la paix entre les hommes. Dès lors, il s'attelle à refonder l'ancienne bibliographie en une véritable science, qu'il nomme documentation, dans la perspective plus globale de la coopération internationale.

Bien qu'ayant eu une influence majeure sur l'émergence de toute une profession, les travaux d'Otlet ont été délaissés par les milieux scientifiques pendant plusieurs décennies, avant de faire l'objet d'une redécouverte progressive (Rayward, 1975 ; Buckland, 1997 ; Van Acker, 2011). Aujourd'hui, une attention nouvelle est portée à son ouvrage majeur, le *Traité de documentation*, notamment dans ses toutes dernières pages, consacrées à une anticipation de l'avenir de la documentation. Dans un passage souvent cité, Otlet imagine un dispositif de consultation que la technologie de l'époque n'est pas tout à fait en mesure de concrétiser et qui nous évoque l'écran d'un ordinateur connecté à une base de données distante (Otlet, 1934, p. 428). Ces quelques lignes d'anticipation devenues célèbres favorisent l'analogie avec nos infrastructures contemporaines. Le Répertoire bibliographique universel, un centre documentaire qui met à profit le réseau téléphonique pour proposer un service informationnel à distance, trouve un lointain successeur dans le Minitel. Quant à Internet et au Web, ils représentent aujourd'hui ce qui se rapproche le plus du projet à échelle mondiale que concevait Otlet.

Le mot documentation a hérité d'Otlet son caractère polysémique. Jean Meyriat en distingue trois acceptions : ensemble de documents, activité constituant cet ensemble et connaissance scientifique associée à cette activité (Meyriat, 1981, pp. 148-149). Nous observons une autre division : la documentation se rapporte à une infrastructure, c'est-à-dire à une organisation de nature à la fois humaine (tour à tour science, profession ou service) et matérielle (supports documentaires, mobilier, lieux de savoir) ou bien à la structure de l'information elle-même. Or comme nous allons le voir, Otlet utilise le réseau comme modèle commun à la structure et à l'infrastructure, unifiant toute la théorie et la pratique de l'organisation des connaissances. Cette logique nourrit une forme d'épistémologie visuelle. Elle fait notamment du graphe une technique privilégiée des humanités numériques et qui peut favoriser l'acquisition d'une culture technique.

2 Documentation et réseau

À l'origine de la proposition d'Otlet se trouve une réflexion sur le livre, dont il juge la forme insatisfaisante au vu des besoins de son époque en matière de communication. Sa critique porte sur deux aspects : la faible densité d'information véritablement utile,

trop souvent noyée dans la glose et le jargon ; la rigidité du format, qui ne permet pas facilement de réutiliser le contenu. C'est au miroir de cette frustration qu'on peut comprendre son intérêt pour la fiche : elle est plus synthétique que le livre et plus facilement manipulable. Cette forme ouvre de nombreuses possibilités, puisqu'une ressource composée de fiches permet plus facilement l'augmentation, l'extraction, la recombinaison, etc. On comprend également son enthousiasme pour les nouvelles technologies d'enregistrement du son et de l'image, qui facilitent la reproduction et la diffusion du texte. C'est pourquoi il propose une définition large du document qui inclut le livre mais aussi ses « substituts », tels que la photographie, le microfilm ou encore le phonogramme. Afin d'organiser ces différents niveaux de granularité et ces formes hétérogènes, Otlet travaille sur des réalisations concrètes bien connues, qui vont influencer les développements de la documentation durant le 20e siècle : la fiche standardisée, associée à un mobilier spécial, et la Classification décimale universelle.

Le *Traité de documentation* constitue l'aboutissement et la synthèse des efforts qu'il déploie pour donner un caractère scientifique à sa démarche. Dans une perspective positiviste influencée par l'atomisme, Otlet traduit l'interchangeabilité du support par l'idée d'une unité informationnelle (le *biblion*) qui peut prendre différentes formes. La mise en rapport de ces unités, dans des ensembles éventuellement complexes, est rendue possible par la classification, qui est conçue comme un codage scientifique des relations entre documents. C'est ainsi que la structure de la documentation se trouve décrite comme un ensemble de nœuds et de liens, c'est-à-dire un réseau. Otlet envisage un temps de publier le *Traité* sous la forme d'une encyclopédie documentaire à la forme révolutionnaire. Malheureusement, les contraintes matérielles qui s'imposent à lui dans les années 1930 ne lui permettent pas de mener à bien ce projet. L'ouvrage paraît sous une forme classique, avec ce sous-titre fort à propos : *Le livre sur le livre*.

Les publications et les archives d'Otlet comportent un nombre important de schémas, depuis les nombreux croquis dont il parsème ses papiers personnels jusqu'aux planches qu'il fait réaliser pour des publications ou des expositions. Dans cette iconographie abondante, Wouter Van Acker distingue certaines métaphores visuelles comme étant centrales dans la pensée d'Otlet, citant comme exemples la sphère, la pyramide et le réseau (Van Acker, 2011, p. 62). De ce dernier, il développe une analyse centrée sur la coopération internationale. S'il fait allusion à une connexion entre les approches universalistes de la connaissance et de la coopération (Van Acker, 2011, p. 96), il n'étudie pas l'application de la métaphore du réseau aux deux approches, seulement à la seconde. Les schémas montrent cependant que la même forme du réseau façonne la vision qu'a Paul Otlet de la connaissance (au sens d'une structure donnée à l'information) et de la coopération (qui constitue une partie de l'infrastructure dédiée à la connaissance), ce qui inscrit sa pensée dans la longue histoire de l'ontologie.

Ainsi que le décrit Umberto Eco, la représentation graphique intervient très tôt dans l'évolution du questionnement ontologique et se manifeste essentiellement sous forme d'arbres (Eco, 2010). En analysant l'arbre de Porphyre (3e siècle), il met en évidence une distinction fondamentale : soit les propriétés qui permettent de fonder l'ontologie représentent la seule façon possible de concevoir le monde et elles façonnent alors un arbre définitif qui produit un mode de connaissance unique ; soit on admet que ces propriétés sont choisies parmi différents groupes qui représentent tous une vision possible et il y a alors autant d'ontologies et d'arbres que de systèmes de pensée. La première approche est limitée : « un arbre à hiérarchie fixe, avec un nombre fini de déterminations, ne sert qu'à classer », il s'apparente au dictionnaire (Eco, 2010, p. 30). Eco montre que la seconde approche s'impose progressivement dans l'histoire de l'ontologie, faisant évoluer ses modèles et ses modes de représentation : « l'arbre éclate en une multitude infinie d'accidents, en un réseau impossible à hiérarchiser de *qualia*. Le dictionnaire finit nécessairement par se dissoudre. Il se fait alors encyclopédie » (Eco, 2010, p. 34).

L'encyclopédie est la forme privilégiée de la documentation otlétienne et constitue également son horizon. La théorie que développe Otlet lui permet en effet de poser les conditions d'un enregistrement total de la connaissance possible ; il en fait le stade ultime du développement de la documentation ou « Hyper-documentation » (Otlet, 1934, p. 429). La possibilité d'inscrire tout ce qui est énoncé socialement apparaît chez Eco dans le concept d'« encyclopédie maximale » : parce qu'il est impossible de déplier un réseau, il propose cette idée comme un horizon régulateur permettant de concevoir « des hiérarchies provisoires, ou des réseaux maniables » (Eco, 2010, p. 74). Un modèle de classification décimale comme celui de Dewey et d'Otlet produit des arbres hiérarchiques ; à l'inverse, la classification à facettes de Ranganathan et les vocabulaires contrôlés utilisés dans le Web sémantique (appelés « ontologies » par abus de langage) forment plutôt un rhizome. En règle générale, il apparaît que la fonction réticulaire de la classification est plus importante que le choix d'un système particulier : c'est elle qui fonde la possibilité d'une *hyperdocumentation*, c'est-à-dire d'un réseau de documents à la fois ouvert, composite, sur différentes échelles et à la structure plus ou moins hiérarchique.

3 Épistémologie visuelle et problématique

La logique hyperdocumentaire est donc le fait de penser la structure de la documentation comme un réseau. Elle s'appuie particulièrement sur un outil visuel bien spécifique : le graphe, c'est-à-dire la représentation graphique d'un modèle abstrait de réseau. Le graphe est un puissant outil de travail interprétatif et réflexif sur la logique hyperdocumentaire – sous certaines conditions.

Les graphes sont à la fois une catégorie générique et une sous-catégorie de représentation : ils incluent par exemple les arbres (qui se distinguent des autres graphes par la présence d'une hiérarchie) mais peuvent être eux-mêmes inclus dans la catégorie des diagrammes. Johanna Drucker situe ces différents outils au sein de l'épistémologie visuelle, qu'elle définit comme « les modes de connaissance qui se présentent et se traitent visuellement » (Drucker, 2014, p. 8) (trad. libre). Elle rappelle que certaines images ont la capacité de générer une connaissance et pas seulement de la représenter. Pour les graphes, elle montre que cette dynamique repose sur deux éléments : l'indexation dans un espace en deux dimensions et une logique combinatoire (Drucker, 2014, pp. 106-107) – nous retrouvons ici les nœuds et les liens du réseau. L'interprétation d'un graphe passe par des parcours de lecture : il est le support d'une circulation intellectuelle dans le réseau. Lorsque le graphe se fait purement illustratif et que sa construction est masquée, cette valeur heuristique est perdue.

Van Acker emploie une terminologie similaire dans son analyse des schémas d'Otlet. Il y distingue trois catégories de métaphores visuelles selon qu'elles remplissent une fonction illustrative, heuristique ou constitutive (Van Acker, 2011,
pp. 58-62). Selon lui, le réseau peut être décrit comme constitutif de la pensée d'Otlet : parce qu'il est intégré profondément à son travail sur une longue période, il devient progressivement une métaphore « morte », au sens où l'analogie entre dessin et idée n'est plus possible que pour les initiés. Otlet utilise le graphe surtout comme symbole, on perd alors ce qui en fait initialement un outil heuristique : l'adéquation totale entre l'objet abstrait (le réseau) et sa représentation concrète (le graphe).

Cette connexion est cruciale, car elle conditionne une certaine réflexivité. Dans son étude des effets de l'écriture sur le langage et la parole, Jack Goody remet en cause la hiérarchie stricte entre compétence et performance défendue par certains linguistes à son époque (Goody, 1979, pp. 60, 143). Le rapport nouveau qu'il établit entre les deux termes est repris par Drucker dans son travail sur l'épistémologie visuelle : « les diagrammes *performent* l'acte de raisonnement, ils ne le représentent pas après coup mais sont le moyen d'en faire fonctionner le processus logique » (Drucker, 2014, p. 115) (trad. libre). Autrement dit, le graphe permet non seulement de générer une connaissance mais d'en expliciter la construction, ce qui nous permet éventuellement de l'analyser. Cette possibilité d'un effet de la performance sur la compétence constitue presque une définition de la réflexivité.

Les humanités numériques nous fournissent un terrain concret pour observer ces idées en action. Les graphes sont les représentations privilégiées des corpus, qui constituent la structure hyperdocumentaire par excellence. Aurélien Berra dit de l'épistémologie des humanités numériques qu'elle concerne avant tout la résolution
pratique de problèmes et les effets de cette résolution – d'où l'expression d'« épistémologie problématique » (Berra, 2014). Celle-ci se puise dans une tradition

particulière de l'anthropologie, de l'histoire et de la philosophie des sciences qui s'intéresse aux médiations techniques telles que le diagramme, la cartographie et la bibliothèque (Dagognet, 1999 ; Jacob, 1992 ; Latour, 2007). Le fil rouge de ces travaux est la boucle établie entre culture et technique à travers différents processus de matérialisation qui sollicitent intellect et outil de façon contigüe. Sa compréhension nécessite ce que nous appelons une pratique des concepts, c'est-à-dire une mise à l'épreuve de la théorie par l'expérimentation.

Or cette pratique ne va pas sans obstacles. Les humanités numériques ont permis la démocratisation d'outils logiciels clé-en-main, ce qui peut entraîner des effets délétères de « boîte noire » (Mounier, 2018, p. 141). Le graphe risque alors d'intervenir alors comme support d'un discours extrêmement fragile sur le plan méthodologique. À cela, on peut opposer une démarche de recherche-conception heuristique. L'Otlétosphère est un exemple (parmi bien d'autres) de cette approche. Cet outil développé au sein du programme ANR HyperOtlet est un prototype de cartographie relationnelle autour de Paul Otlet[1]. Les données (entités et relations) sont stockées dans un format tabulaire à partir duquel le graphe est généré. La circulation constante entre ces deux niveaux est bénéfique pour comprendre qu'il n'est de lien qui ne soit construit. L'outil permet de tester la base de connaissances et le modèle *a priori* des relations. Ce type de dispositif permet de se confronter à la structure hyperdocumentaire du corpus.

4 Conclusion

La forme que conçoit Paul Otlet pour la documentation au début du 20e siècle représente une innovation considérable dans l'organisation des connaissances, qui structure toujours nos systèmes d'information aujourd'hui. C'est pourquoi il est crucial d'en saisir la logique, que nous avons qualifiée d'hyperdocumentaire. Celle-ci nous permet de mieux comprendre le lien entre structuration et visualisation de l'information dans un champ comme les humanités numériques, au sein duquel l'usage des graphes est très répandu. Cette logique appelle une épistémologie à la fois visuelle et problématique, c'est-à-dire qui étudie la visualisation dans la perspective de la réflexivité. Elle nous invite à nous intéresser à des outils qui ne sont pas seulement informatifs mais formateurs.

1 http://hyperotlet.huma-num.fr/otletosphere.html

Références

Berra, A. (2014). L'épistémologie problématique des humanités numériques. *Philologie à venir*. [Billet]. https://philologia.hypotheses.org/1461

Buckland, M. K. (1997). What is a « document »? *Journal of the American Society for Information Science*, *48*(9), 804-809. DOI : 10.1002/(SICI)1097-4571(199709)48:9<804::AID-ASI5>3.0.CO;2-V

Dagognet, F. (1999). *Les outils de la réflexion : épistémologie*. Le Plessis-Robinson : Institut Synthélabo pour le progrès de la connaissance.

Drucker, J. (2014). *Graphesis: Visual Forms of Knowledge Production*. Cambridge, Massachusetts : Harvard University Press.

Eco, U. (2010). *De l'arbre au labyrinthe. Études historiques sur le signe et l'interprétation*. Paris : Grasset.

Goody, J. (1979). *La Raison graphique : la domestication de la pensée sauvage*. Paris : Les Editions de Minuit.

Jacob, C. (1992). *L'empire des cartes. Approche théorique de la cartographie à travers l'histoire*. Paris : Albin Michel.

Latour, B. (2007). Pensée retenue, pensée distribuée. Dans C. Jacob (Éd.), *Lieux de savoir* (pp. 605-615). Paris : Albin Michel.

Meyriat, J. (1981). Document, documentation, documentologie. Dans V. Couzinet (Éd.), *Jean Meyriat, théoricien et praticien de l'information-documentation* (rééd., 2001). Paris : ADBS Éditions.

Mounier, P. (2018). *Les humanités numériques : une histoire critique*. Paris : Éditions de la Maison des sciences de l'homme. http://books.openedition.org/editionsmsh/12006

Otlet, P. (1934). *Traité de documentation. Le livre sur le livre* (3e éd.). Bruxelles : Les Impressions nouvelles (2015).

Rayward, W. B. (1975). *The universe of information: The work of Paul Otlet for documentation and international organisation*. Moscow : All-Union Institute for Scientific and Technical Information.

Van Acker, W. (2011). *Universalism as utopia : A historical study of the schemes and schemas of Paul Otlet (1868-1944)*. dissertation. Ghent University. http://hdl.handle.net/1854/LU-1942681

1.4 Démarche de recueil de pratiques en vue de l'émergence d'un modèle d'élaboration didactique d'un cours de thermodynamique dans l'enseignement supérieur.

Pierre-Bernard Toubol - (présentation)

Démarche de recueil de pratiques en vue de l'émergence d'un modèle d'élaboration didactique d'un cours de thermodynamique dans l'enseignement supérieur.

Pierre-Bernard TOUBOL

COMUE Languedoc-Roussillon Universités
8, rue de l'école normale CS 78290 - 34197 Montpellier cedex 5
pierre-bernard.toubol@lr-universites.fr

Résumé Dans un projet structuré autour d'un objectif de formation, un enseignant identifie puis organise les concepts propres à la discipline qu'il veut enseigner. Considérant qu'il ne suffit pas de bien connaître une matière pour l'enseigner efficacement et qu'enseigner plus efficacement suppose une meilleure connaissance des trois pôles essentiels mis en jeu ainsi que leurs relations mutuelles : le savoir, l'enseignant et l'apprenant (Viennot 1989), nous souhaitons bâtir un modèle permettant de structurer les connaissances disciplinaires destinées à alimenter un outil d'aide à la production d'un parcours de formation. D'un point de vue méthodologique nous nous proposons de construire une chaîne de production de connaissances s'appuyant d'une part sur une analyse du discours issue d'entretiens avec différents acteurs (enseignants, ingénieurs pédagogiques, apprenants) et d'autre part sur l'analyse de parcours pédagogiques déjà disponibles.

Mots-clés: web sémantique, environnement informatisé pour l'apprentissage humain, traitement automatique du langage naturel, réseaux lexicaux

1 État de l'art et objectifs de la recherche

Dans le paysage actuel de l'enseignement supérieur français, on assiste depuis plusieurs années à une massification des enseignements. En réponse à un besoin croissant de personnalisation des apprentissages en fonction des acteurs (apprenants et enseignants), des contextes et des objectifs de formation pédagogique, les systèmes de gestion numérique d'apprentissage en tant qu'environnement informatisés pour l'apprentissage humain (EIAH) requièrent un accompagnement didactique des enseignants dans la création des scénarisations de leur enseignements. C'est dans un contexte scientifique pluridisciplinaire convoquant les sciences de l'éducation, la didactique des disciplines, l'ingénierie des connaissances et l'informatique que notre démarche s'inscrit. Nous souhaitons ainsi montrer que la *didactique* en tant que modélisation d'une connaissance en vue de son apprentissage, est une démarche pouvant s'appuyer sur une logique *épistémologique* que l'on pourrait décrire et générer à l'aide d'une méthode d'analyse sémantique automatisée.

Le concept d'épistémologie est choisi ici, dans sa *participation* à l'élaboration d'une science du langage scientifique par la reconnaissance d'unité des formes de la connaissance dans leur rapport avec la science ("sens commun" vs. connaissances scientifiques). Par ailleurs, les connaissances sont en perpétuelles croissance : c'est un problème central pour l'épistémologie. Étudier son évolution au regard des connaissances scientifiques était d'ailleurs la solution proposée par Popper "«Le problème central de l'épistémologie a toujours été et reste le problème de la croissance des connaissances, et la meilleure façon d'étudier cette dernière est d'étudier la croissance de la connaissance scientifique" (cité par Dessus (2007).)

Le choix de la thermodynamique nous sert de fil rouge. Cette science s'appuie sur un champ conceptuel et lexical très large faisant appel des disciplines multiples. Elle permet ainsi des incursions dans des domaines scientifiques aussi différents que la chimie, la physique, les mathématiques mais aussi l'histoire (en raison de sa récente émergence en tant que science[1]). La thermodynamique est une discipline cherchant à comprendre les caractéristiques macroscopiques de systèmes caractérisés par des variables comme la pression, le volume ou la température. Depuis le 19ème siècle, la thermodynamique statistique cherche à comprendre les effets des états microscopiques sur le système macroscopique.

Il est usuel de présenter la didactique de la thermodynamique comme un champ scientifique au carrefour de divers autres champs déjà cités : mathématiques, épistémologie, sciences de l'ingénieur, chimie.... et, tout en soulignant le rôle que peuvent jouer (ou ont joué) ces sciences dans son développement, d'insister sur le fait que la problématique didactique conduit à remanier plus ou moins profondément les outils, conceptuels ou méthodologiques, que la recherche leur emprunte.

Dans la suite de notre travail on s'appuiera sur les relations entre épistémologie et didactique et plus précisément les besoins épistémologiques en didactique. Nous définirons ainsi les besoins formulables en terme de connaissance des processus par lesquels les concepts utilisés en thermodynamique ont été formés et développé (épistémologie historique) et plus généralement de connaissances caractéristiques structurant cette discipline.

Pascal Charbonnat Charbonnat et al. (2014) dans son approche de "la phylogénie des concepts savants", définit sa démarche en traitant non pas des objets culturels en général mais d'un objet spécifique attaché aux énoncés de connaissance : le concept. "Un concept est l'ensemble des rapports entre une forme lexicale stable qui se répète durant un intervalle de temps fini par un ou plusieurs locuteurs, et des formes lexicales cooccurrentes qui accompagnent de manière variable cette forme stable, et des événements non lexicaux (sociaux, cérébraux, biologiques,

[1] "l'histoire des sciences elle-même peut être prise comme contenu enseigné dans le cadre de l'éducation scientifique.(Martinand, 1993)

etc.)". Ainsi, il fait apparaître au cœur de sa définition la notion de contexte que nous chercherons à modéliser par les relations internes aux corpus textuels. Un mot n'a de sens que dans un contexte. Selon Wittgenstein "l'usage guide le sens. Le point de départ de sa pensée consiste à identifier le sens des mots à leur usage Wittgenstein, Granger (2002). Les mots ont des sens complexes et variés qui dépendent fortement du contexte. Il n'y a pas de conceptualisation à priori qui existerait en dehors d'une pratique sociale ou d'une situation donnée.

Cette analyse est complétée au niveau de l'apprenant par la prise en compte de de ses propres "conceptions". Selon Artigue (1989), le terme de conception décrit la mise en évidence de la pluralité des points de vue possibles sur un même objet, la différentiation des représentations et des modes de traitement qui lui sont associés , et la mise en évidence de leur adaptation plus ou moins bonne à la résolution de problème.

Par ailleurs le terme de conception aide le didacticien à différentier le savoir que l'enseignant veut transmettre et les connaissances effectivement construites par l'apprenant. "Le terme de «conception» apparaît donc comme un outil pour la construction d'un concept permettant la modélisation de l'élève en tant qu'agent apprenant" Balacheff, Margolinas (2003)

"La conception est un processus d'une activité de construction mentale du réel. Il y a conception ou activité de conceptualisation lorsqu'un objet ou les éléments d'une série d'objets se trouvent :

- figurés
- exprimés
- traduits

et qu'une correspondance se traduit par des relations qui relient entre eux les éléments qui font l'objet de la représentation". Giordan, De Vecchi (1987)

Afin de prendre en compte l'apprenant dans notre démarche, il nous est apparu nécessaire de choisir un modèle d'apprenant qui soit suffisamment opérationnel pour notre travail. Nous avons retenu la taxonomie de Bloom[2]. Cette notion de contexte a donnée lieu à une variété très importante de définitions analysée dans Bazire, Brézillon (2005). Parmi celles-ci, la définition opérationnelle que nous avons retenue est constituée de l'ensemble des circonstances qui encadrent un événement ou un objet : le contexte se définit ainsi comme ce qui manque à un objet donné pour qu'un utilisateur puisse s'en construire une représentation correcte. Nous avons choisit d'extraire puis d'exprimer formellement les relations entre les différentes formes lexicales impliquées et les relations entre "acteurs" de l'apprentissage que nous identifierons par des entretiens avec ces acteurs. Cette

[2] La raison de ce choix est explicitée dans un travail externe à cette communication.

démarche permettra de mettre en évidence puis d'analyser différentes situation d'obstacles épistémologique rencontrés et décrits par exemple en mathématique par Brousseau (1976), en physiques par Viennot (1996) et en thermodynamique par Rozier (1988).

Plusieurs cadres théoriques ont été signalés dans la littérature parmi lesquels :

- La théorie des situations didactiques Brousseau, Balacheff (1998) appliquée aux mathématiques : le savoir (connaissance de référence) est représenté par le problème où il intervient. Ainsi, à un savoir donné correspond un ensemble de situations déterminé par les variables didactiques
- La théorie APOS Dubinsky, Mcdonald (2002) centrée sur la compréhension de l'apprentissage des concepts chez un individu. Cette théorie considère que la manière dont un savoir est appris dépend de la structure de ce savoir. Ainsi la première étape d'une recherche fondée sur ce cadre théorique, doit être une analyse du savoir en jeu, permettant d'en proposer une décomposition génétique qui met en évidence les différentes conceptions que les étudiants peuvent avoir.
- L'approche onto-sémiotique Godino (2005) s'intéresse aux « systèmes de pratiques (opérationnelles et discursives) mises en œuvre par les personnes impliquées dans certains types de situations problématiques. Sont définit six types d'objets : un langage ; des situations ; des concepts ; des propositions, propriétés ou attributs ; des procédures et des arguments et enfin des relations de dépendances entre ces objets au travers de leurs fonctions sémiotiques
- La Théorie Anthropologique du Didactique (TAD) construite comme une généralisation de la théorie de la transposition didactique Chevallard Yves (1985), il s'agit d'une théorie vivante et en construction. Pour appréhender la TAD il y a trois notions fondamentales, incarnées par ce que Chevallard appelle des termes primitifs : « les objets O , les personnes X , les institutions I ».

"Au-delà de sa légitimité sociale et culturelle, chaque savoir enseigné se justifie en droit par sa correspondance à un savoir extra-scolaire - un savoir savant - qui le précède et qui le fonde culturellement et scientifiquement... Or, nécessairement, dans ce passage du savoir savant au savoir enseigné, un écart se creuse, une transposition s'opère, qui adapte, remanie, et parfois dénature les éléments empruntés à la sphère des pratiques savantes de la connaissance." (Chevallard)

La différence entre savoir et connaissance est l'un des fondements essentiels de la théorie des situations didactiques (TSD) selon Guy Brousseau. "On considère que le système éducatif peut être modélisé par les relations entre un système "enseignant" et un système "enseigné", et une situation d'enseignement par les relations entre les sous-systèmes : le système didactique, le système apprenant, le contexte." Brousseau, Balacheff (1998)

L'enquête entreprise a donc pour objectif la mise en évidence des pratiques didactiques que les enseignants intervenant dans les différents cycles universitaires mettent en œuvre. Elle est en cours de réalisation. En effet, c'est au travers de l'analyse des corpus construits à partir d'entretiens semi-directifs lors de la première partie de la recherche entreprise et d'enquête par observation concernant notamment une mise en situation d'élèves ingénieurs de l'école de chimie de Montpellier (évaluation de prérequis) que nous nous proposons de faire émerger une philogénèse et une ontogénèse des "concepts savants" de thermodynamique. Cette analyse de corpus sera conduite selon une méthode de théorisation ancrée (fondée par Glaser et al. (1968) et reprise par Corbin, Strauss (1990) puis systématisée par Lejeune (2016)). Nous procédons à une première réduction du modèle TSD en nous restreignant aux liens entre système didactique et système apprenant et en choisissant comme observable la notion de conception. L'avantage de cette méthode par rapport à une méthode numérique statistique est ici de nous permettre d'orienter et d'introduire du sens dans l'analyse.

Notre choix du modèle de référence de travail s'appuie sur la modélisation de la connaissance dans la Théorie Anthropologique du Didactique définie par Chaachoua (2010). Toute activité humaine met en œuvre une organisation que Chevallard (id. cit.) note $T/\tau/\theta/\Theta$ et qu'il nomme praxéologie, ou organisation praxéologique.

- T/τ étant la pratique – ou encore le savoir- faire ;
- θ/Θ le logos – ou encore le savoir.

On s'appuiera sur la praxéologie didactique avec comme type de tâche T les tâches d'étude propre à la discipline choisie. Par ailleurs, une extraction des relations sémantiques sera réalisée sur un ensemble de corpus didactiques. Cette capture de relations faisant appel à des techniques de TALN constitue un moyen privilégié nécessaire à l'alimentation d'une base de connaissance structurée dans un réseau lexical. Ce type de réseau est une des formes de base de connaissances plus à même de représenter une phylogénie de concepts du type de celles décrite par Charbonnat et al. (2014) qu'une ontologie. Il est possible en outre d'y intégrer leur ontogénèse.

En effet, les ontologies de domaine issues de bases de connaissances spécifiques ont une valeur certaine mais limitées dans leur portée. Il manque des méta-informations sur la force d'association (les poids) et les annotations (fréquences, pertinence). Le projet JDM : réseau lexical généraliste "Jeux de mots" Lafourcade, Joubert (2008) intègre une méthode d'apprentissage sans fin permettant de générer des règles de façon inductive à partir des faits présents. Ces règles permettent de densifier le réseau et d'en réduire les silences. Afin de minimiser l'émergence de relations qui pourraient être erronées.

Une fois instruite cette base de donnée, complétée par des éléments de phylogénèse issu de l'analyse qualitative des entretiens permettra de constituer un modèle destiné à produire un réseau conceptuel d'apprentissage. Chaque concept/catégorie extrait par théorisation ancrée est projeté dans le réseau lexical. Ce réseau combinera poids et annotations sur des relations typées entre des termes et des concepts. Il sera à l'origine de référentiels devant aboutir à la proposition de parcours de formation[3].

2 Méthode et entretiens déjà réalisés

Plutôt que de s'appuyer sur une théorisation "toute faite" de la diffusion des connaissances, savoirs et pratiques, nous cherchons à bâtir un modèle opérant à partir des pratiques de terrain. (analyse d'entretiens et de corpus)

La structuration de nos guides d'entretien (s'adaptant aux deux types d'interlocuteurs : enseignants ou apprenants) est dynamique et évolutive. Leur adaptation se déroule de manière incrémentale selon les catégories émergentes d'une analyse qualitative ancrée. L'analyse de discours issus des entretiens eux-mêmes est destinée à recueillir et questionner les pratiques Pélissier (2017) et les expériences des acteurs de la formation concernant la création d'un parcours de formation pour la discipline choisie.

Nous avons fait le choix de privilégier une approche "Système d'Information". Par exemple, pour l'analyse du processus de l'aide à la transmission de connaissances, celui-ci est vu comme un processus[4] interne au Système d'Information (processus de support au sens de Porter, Millar (1985). Ainsi, le processus d'aide est décrit grâce à un formalisme de description et de documentation concernant :

- les acteurs
- les étapes
- les produits
- les indicateurs de qualité

Nous définissons le processus de l'aide par son périmètre (son contexte). Cet ensemble d'activités reliées entre elles, réalisées par des acteurs et concourant à atteindre un objectif fixé s'inscrit dans une description de l'enchaînement des activités nécessaires à la production des résultats. Qui fait quoi et comment :

[3] Nous désignons ici "parcours de formation" une arborescence didactique de concepts savants destinée à baliser un apprentissage en vue de d'atteindre un objectif didactique de formation (ex : quelle est la hiérarchie et l'ordre des concepts à maîtriser pour atteindre tel objectif de formation)

[4] Nous définirons un processus comme un ensemble d'activités reliées entre elles, réalisées par des acteurs et concourant à atteindre un objectif fixé. Il décrit l'enchaînement des activités nécessaires à la production des résultats.

- les acteurs qui interviennent ;
- les actes réalisés (transformation d'information) ;
- l'ordre des séquences (mise en évidence des frontières entre acteurs) et leur description incluant les outils de traitement.

Trois paradigmes (vues du système d'information) sont alors posés : l'organisation (acteurs), les informations (concepts et inter-relation) et les échanges (flux d'information).

La vue "échanges" est une vue extérieure au processus. Elle décrit les échanges d'information entre les interlocuteurs à qui ce processus destine les résultats de son action ainsi que tous les acteurs dont il dépend pour produire ces résultats.

La vue "organisationnelle/fonctionnelle" est une vue interne. Elle décrit les acteurs (ou groupe d'acteurs) physiques ou organisationnels qui collaborent à l'atteinte des résultats (objectifs). Ces acteurs sont illustrés par leurs compétences (domaine de responsabilité, pouvoir de décision). Elle met en évidence la façon dont l'organisation prend en charge ses missions.

La vue "informationnelle" considère l'ensemble des informations manipulées (les concepts et leurs inter-relations) qui constituent un micro-univers (éléments du contexte) : la représentation restreinte du monde dont s'est dotée l'organisation pour accomplir ses missions. La démarche consiste donc à identifier, extraire, et modéliser l'information. L'information n'est pas une information structurée : elle a donc besoin d'être modélisée, formalisée afin de simplifier son analyse. On cherche à en décrire son cycle de vie, depuis sa production jusqu'à sa délivrance.

Le recueil et l'identification des informations récoltées et articulées les unes par rapport aux autres constitue les premiers éléments de la phylogénie des concepts identifiés. L'alimentation du guide de codage (au sens de Lejeune) issu du recueil s'effectue à la "volée". Il s'appuie sur les deux niveaux de catégorie proposé dans le cadre d'un outil d'encodage RQDA[5] Dunn (2012). Une "condensation" des catégories d'analyse à partir des étiquettes assignées à des portions de corpus illustre une typologie de l'aide réellement mise en œuvre. Ce codage est lié à la vue organisationnelle/fonctionnelle. Elle concerne ici :

- l'identification des acteurs
- les relations entre acteurs
- l'identification des actions
- le séquençage des actions

[5] RDQA est un paquet R pour l'analyse qualitative des données, une application logicielle d'analyse qualitative gratuite (libre comme la liberté) (licence BSD).De plus, il s'intègre parfaitement avec R, ce qui signifie que l'analyse statistique du codage est possible, et que les fonctions de manipulation et d'analyse des données peuvent être facilement étendues en écrivant des fonctions R. Dans une certaine mesure, RQDA et R constituent une plate-forme intégrée pour l'analyse quantitative et qualitative des données.

Le codage lié à la vue informationnelle permet l'identification des "objets traités".

Une chaîne de traitement numérique résolument basée sur des outils "open-source" et ouverts permet des traitements multiples au travers :

- d'outils dédiés (RQDA, Tropes, Iramuteq, TXM) ;
- de langages orientés vers le traitement de données et l'analyse statistique (R ou Python).

Le protocole est définit de la manière suivante :

- La captation : enregistrement audio
- La transcription : (Transcriber 1.5.1) permet :
  - Une identification standardisée des locuteurs (ingp, prof, etud, pbt)
  - Une utilisation restreinte des balises (entités nommées, prononciation, commentaires)
- Des exports textuels permettant les corrections orthographiques (coquilles ou grammaire)
- L'intégration dans le CAQDAS (Computer Aided Qualitative Data Analysis System) retenu : RQDA permettant la création d'une base de données relationnelle. Des interrogations de type SQL sont alors réalisées par croisement de donnée ou traitement par langage R ou Python.

Le corpus actuel est constitué par la transcription de 21 entretiens (présentiels ou à distance) avec des acteurs de l'enseignement supérieur :

- des enseignants et des chercheurs (universités françaises et canadiennes)
- des ingénieurs pédagogiques (université françaises, écoles et grands établissements)
- des étudiants (Québec)

Notre guide de codage s'appuie donc sur les deux niveaux de catégorie proposés dans le cadre de RQDA. La "condensation" des catégories d'analyse à partir des étiquettes assignées à des portions de corpus illustre une typologie du processus mis en œuvre.

A titre d'exemple une première catégorisation (provisoire) d'aide s'exprime de la manière suivante :

Catégories
Mise en confiance et empathie Exploration : susciter une situation, émettre des hypothèses, faire préciser un mot ou un concept. Nature de la construction à partir de la didactique de la discipline : poser et hiérarchiser les concepts, mettre à jour les contenus.

Catégories
Définition du champ d'intervention dans le dispositif pédagogique proposé : identifier l'accompagnement dans ses objectifs identifier le paramètre institutionnel. Relation entre l'apprenant, son apprentissage et la personne aidée : définir les prérequis, définir les objectifs, gérer les profils d'apprenant. Appui à la pratique pédagogique : positionner une action dans une démarche d'apprentissage, qualifier et évaluer une pratique, relier un propos à une théorie de l'apprentissage.

Ces catégories constituent des éléments clés pour :

- une base à l'alimentation du guide visant à faire expliciter par la personne interviewée la nature des activités qu'elle mène en tant qu'enseignant produisant seul un cours ou en tant qu'ingénieur pédagogique aidant un enseignant à structurer son cours.
- une mise en évidence d'éventuelles ruptures dans la chaîne de l'organisation des connaissances en vue de les transmettre.
- alimenter la démarche de production d'un outil informatique intégré destiné aux enseignants pour les aider à structurer les connaissances à transmettre et aux ingénieurs pédagogiques pour les aider dans leur travail d'aide à l'enseignant.

Une autre modalité d'enquête est mise en œuvre lors d'une session d'évaluation de prérequis auprès d'élève ingénieurs. Elle entrera en tant que telle dans le corpus.

Références bibliographiques

ARTIGUE, Michèle, 1989. *Épistémologie et Didactique.* [en ligne]. S.l. : IREM. Cahier de DIDIREM, 3. Disponible à l'adresse : http://docs.irem.univ-paris-diderot.fr/up/publications/IPS97066.pdf.

BALACHEFF, Nicolas et MARGOLINAS, Claire, 2003. cK Modèle de Connaissances Pour Le Calcul de Situations Didactiques. In : *XII° École d'été de Didactique Des Mathématiques* [en ligne]. S.l. : La Pensée Sauvage éditions. 2003. p. 1-32. [Consulté le 20 juin 2017]. Disponible à l'adresse : https://hal.archives-ouvertes.fr/hal-01139408/.

BAZIRE, Mary et BRÉZILLON, Patrick, 2005. *Understanding Context Before Using It.* S.l.

BROUSSEAU, Guy, 1976. *Les Obstacles Épistémologiques et Les Problèmes En Mathématiques.* S.l. : Louvain-la-neuve.

BROUSSEAU, Guy et BALACHEFF, Nicolas, 1998. *Théorie Des Situations Didactiques : Didactique Des Mathématiques 1970-1990*. S.l. : La pensée sauvage Grenoble. ISBN 2-85919-134-8.

CHAACHOUA, Hamid, 2010. *La Praxéologie Comme Modèle Didactique Pour La Problématique EIAH. Etude de Cas : La Modélisation Des Connaissances Des Élèves*. HDR. S.l. : Université de Grenoble.

CHARBONNAT, Pascal, BEN HAMED, Mahé et LECOINTRE, Guillaume, 2014. *Apparenter la pensée ? Vers une phylogénie des concepts savants.* [en ligne]. S.l. : Editions Matériologiques. [Consulté le 2 septembre 2019]. ISBN 978-2-919694-56-3. Disponible à l'adresse : http://www.cairn.info/apparenter-la-pensee--9782919694563.htm.

CHEVALLARD YVES, 1985. *La Transposition Didactique : Du Savoir Savant Au Savoir Enseigné*. Grenoble : Grenoble : Éd. la Pensée sauvage, DL 1985.

CORBIN, Juliet M. et STRAUSS, Anselm, 1990. Grounded Theory Research : Procedures, Canons, and Evaluative Criteria. In : *Qualitative sociology*. 1990. Vol. 13, n° 1, p. 3-21.

DESSUS, Philippe, 2007. La Théorie de La Connaissance de Popper et Ses Implications Pour l'éducation. In : *Penser l'éducation*. 2007. Vol. 21, p. 19-31.

DUBINSKY, Ed et MCDONALD, Michael A., 2002. APOS : A Constructivist Theory of Learning in Undergraduate Mathematics Education Research. In : HOLTON, Derek, ARTIGUE, Michèle, KIRCHGRÄBER, Urs, HILLEL, Joel, NISS, Mogens et SCHOENFELD, Alan (éd.), *The Teaching and Learning of Mathematics at University Level* [en ligne]. Dordrecht : Kluwer Academic Publishers. p. 275-282. [Consulté le 15 septembre 2019]. ISBN 978-0-7923-7191-5. Disponible à l'adresse : http://link.springer.com/10.1007/0-306-47231-7_25.

DUNN, Lindsey, 2012. *R QUALITATIVE DATA ANALYSIS (RQDA) PACKAGE : A FREE QUALITATIVE DATA ANALYSIS TOOL Learn How to Import and Work with Interview Data in R* [en ligne]. [Consulté le 26 mars 2019]. Disponible à l'adresse : http://comm.eval.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=4bc47744-23e2-445f-90e1-f1ad8b6ec30b&forceDialog=0.

GIORDAN, André et DE VECCHI, Gérard, 1987. Les Origines Du Savoir. In : *Neuchâtel, Delachaux et Niestlé*. 1987.

GLASER, Barney G., STRAUSS, Anselm L. et STRUTZEL, Elizabeth, 1968. The Discovery of Grounded Theory ; Strategies for Qualitative Research. In : *Nursing research*. 1968. Vol. 17, n° 4, p. 364.

GODINO, Juan, 2005. The Onto-Semiotic Approach to Research in Mathematics Education. In : *Yupana* [en ligne]. 20 novembre 2005. n° 2, p. 43-60. [Consulté le 15 septembre 2019]. DOI 10.14409/yu.v1i2.244. Disponible à l'adresse : https://bibliotecavirtual.unl.edu.ar/publicaciones/index.php/Yupana/article/view/244.

LAFOURCADE, Mathieu et JOUBERT, Alain, 2008. Jeuxdemots : Un Prototype Ludique Pour l'émergence de Relations Entre Termes. In : *JADT'08 : Journées Internationales d'Analyse Statistiques Des Données Textuelles* [en ligne]. S.l. : s.n. 2008. p. 657-666. [Consulté le 9 mars 2016]. Disponible à l'adresse : https://hal.archives-ouvertes.fr/lirmm-00358848/.

LEJEUNE, Christophe, 2016. Logiciels d'analyse de données qualitatives ou d'analyse qualitative ? In :. septembre 2016. p. 13.

MARTINAND, Jean-Louis, 1993. Histoire et Didactique de La Physique et de La Chimie : Quelles Relations ? In : *Didaskalia (Paris)*. 1993.

PÉLISSIER, Chrysta, 2017. *L'aide En Question : Un Composant Majeur Dans l'éducation et Les Humanités Numériques*. Habilitation à diriger des recherches. S.l. : Paul Valéry, Montpellier 3.

PORTER, Michael E. et MILLAR, Victor E., 1985. How Information Gives You Competitive Advantage. In : *Harvard Business Review* [en ligne]. 1985. Vol. 63, n° 4, p. 149-160. Disponible à l'adresse : http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=8500002422&lang=fr&site=ehost-live.

ROZIER, Sylvie, 1988. *Le Raisonnement Linéaire Causal En Thermodynamique Classique Élémentaire*. S.l. : Université Denis Diderot Paris VII.

VIENNOT, Laurence, 1989. L'enseignement des Sciences Physiques, objet de recherche. In : *Bulletin de l'Union des Physiciens* [en ligne]. septembre 1989. n° 716, p. 12. Disponible à l'adresse : http://bupdoc.udppc.asso.fr/index.php.

VIENNOT, Laurence, 1996. *Raisonner En Physique : La Part Du Sens Commun*. S.l. : De Boeck Supérieur. ISBN 2-8041-2236-0.

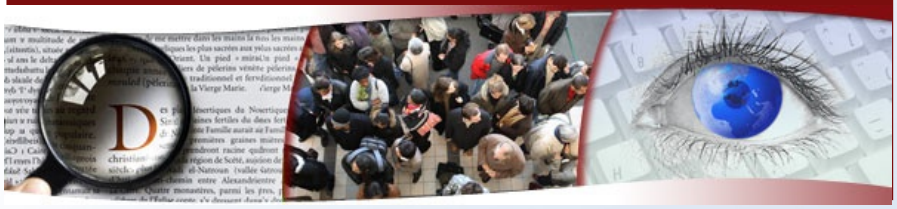
WITTGENSTEIN, Ludwig et GRANGER, Gilles-Gaston, 2002. *Tractatus logico-philosophicus*. Paris : Gallimard. ISBN 978-2-07-075864-7.

PRAXILING

LIRMM

cnrs

Praxiling UMR 5267 -
Université de Montpellier 3 -
CNRS



Colloque Jeunes Chercheurs PRAXILING 2019

«Constitution assistée d'un parcours de formation s'appuyant sur les outils du web sémantique et l'analyse de corpus textuels »

Présentation des travaux

Pierre-Bernard TOUBOL
Ingénieur de recherche
Doctorant
Humanités Numériques

Sous la direction de :

Chrysta Pelissier
PRAXILING - Équipe 2
et :
Mathieu Lafourcade
LIRMM – Équipe TEXTE

CJCPRAX - novembre 2019 1

Menu

A. État de l'art et objectifs de la recherche et substrats théoriques

B. Démarche, méthodes et premiers résultats

C. Participations et communications

D. Planning prévisionnel

Positionnement de la recherche

- Massification des enseignements dans le supérieur.
- Besoin de personnalisation des apprentissages fonction des acteurs (apprenants et enseignants), des contextes et des objectifs (de formation et pédagogiques).
- Systèmes de gestion numérique d'apprentissage (EIAH) nécessitant un accompagnement des enseignants dans la création de scénarisations de leur enseignement.

Objectif de la recherche

- Un objectif :
  - bâtir une méthode, et un modèle permettant de structurer les connaissances disciplinaires destinées à alimenter un outil d'aide à la production d'un parcours de formation.

La didactique en tant que modélisation d'une connaissance en vue de son apprentissage, est une démarche pouvant s'appuyer sur une logique *épistémologique* que l'on pourrait décrire et générer à l'aide d'une méthode d'analyse sémantiques automatisée.

Épistémologie et langage scientifique

- Épistémologie en tant que concept participant à l'élaboration du langage scientifique par la reconnaissance d'unité des formes dans la connaissance dans leur rapport avec la science
- Les connaissances sont en perpétuelles croissance : c'est un problème central pour l'épistémologie...

  → nécessité de la prise en compte de leur évolution

Choix du champ disciplinaire

La thermodynamique

- Champs conceptuel et lexical très large faisant appel à des disciplines multiples : la chimie, la physique, les mathématiques et... l'histoire en raison de sa récente émergence en tant que science
- Il est d'usage de présenter sa didactique comme au carrefour de toutes ces champs disciplinaires qui ont participé à son développement.
- La problématique didactique conduit a remanier les outils conceptuels ou méthodologiques que la recherche leur emprunte (S. Rozier 1988)

"Phylogénie" des concepts savants

“Un concept est l’ensemble des rapports entre une forme lexicale stable qui se répète durant un intervalle de temps fini par un ou plusieurs locuteurs, des formes lexicales cooccurrentes qui accompagnent de manière variable cette forme stable, et des événements non lexicaux (sociaux, cérébraux, biologiques, etc.)." (P. Charbonnat 2014)

- Charbonnat emprunte à la biologie la notion de phylogénie qui correspond à l'étude des liens existant entre concepts apparentés.
- Il fait apparaître au cœur de sa définition la notion de contexte que nous chercherons à modéliser par les relations internes aux corpus textuels
- Pour bâtir ces contextes, nous chercherons à exprimer formellement ces relations entre les différentes formes lexicales et par les relations entre “acteurs” de l’apprentissage

  Des analyses approchant cet objectif (mise en évidence des "obstacles épistémologiques") ont été développées par G. Brousseau en mathématiques ou L. Viennot en physique ou encore S. Rozier en thermodynamique.

Objectif des enquêtes mises en œuvre

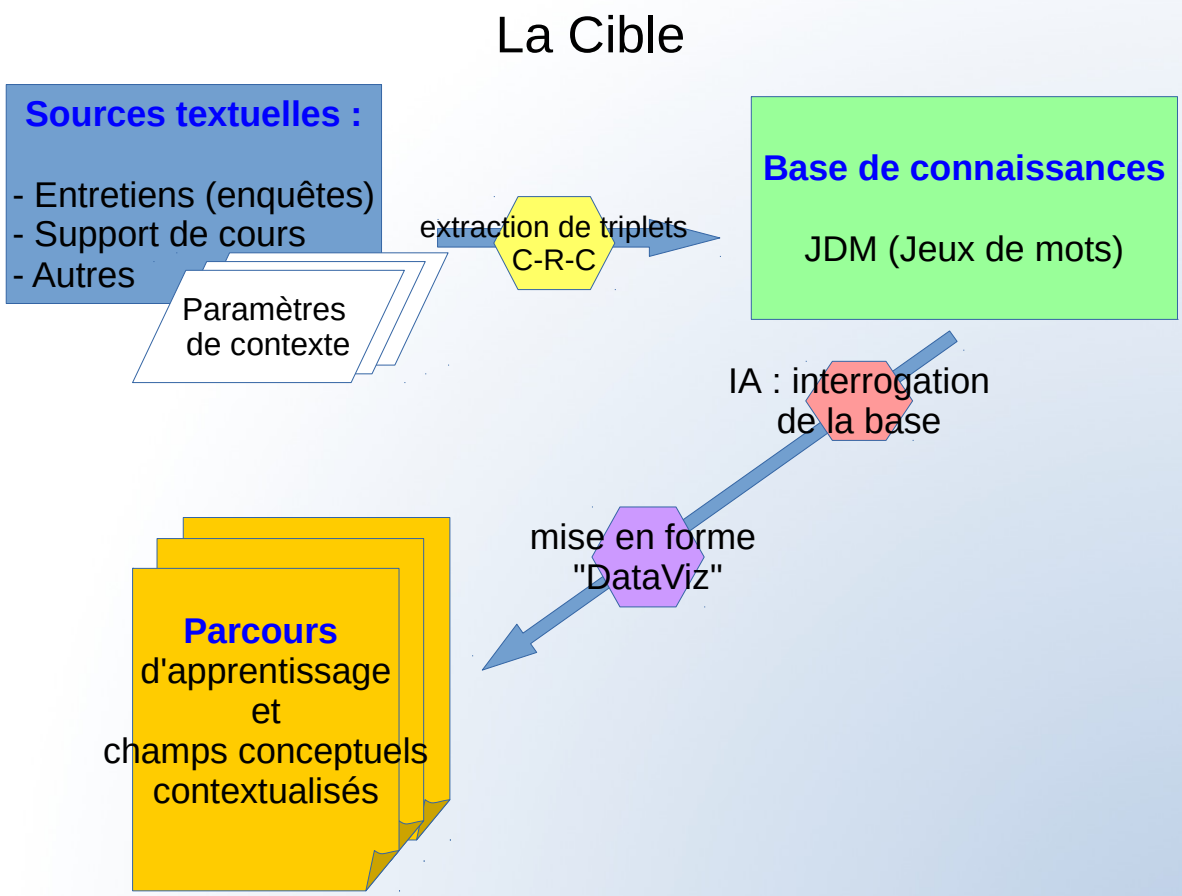
- Mettre en évidence et analyser des pratiques didactiques des enseignants du supérieur
  - entretiens semi-directifs
  - enquêtes par observation
- Faire émerger une "phylogénèse" des concepts savants de thermodynamique par une analyse de corpus conduite selon une méthode de "théorisation ancrée" (Glaser et al. 1968, Strauss et Corbin 1990, Lejeune 2016)

Corpus didactiques : identification et capture des relations sémantiques

- Après sélection d'un ensemble de support de cours (en accès libre ou bénéficiant d'une autorisation explicite)
- Capture des relations sémantiques en faisant appel à des techniques de TALN.
- Alimentation de la base de connaissance structurée dans le réseau lexical JDM (Lafourcade 2008) : plus approprié pour modéliser une phylogénie ainsi qu'une ontogénèse (évolution dans le temps) de concepts qu'une simple ontologie.

Réseau lexical vs ontologie de domaine

- Les ontologies de domaine issues de bases de connaissances spécifiques ont une valeur certaine mais sont limitées dans leur portée, spécialisé et dont l'agilité est limitée.
- JDM apporte des méta-informations
  - la force d'association (poids)
  - des annotations (fréquence et pertinence)
- réseau lexical généraliste "Jeux de mots" (Lafourcade 2008) intègre une méthode d'apprentissage sans fin.


La Cible
Sources textuelles :
- Entretiens (enquêtes)
- Support de cours
- Autres
Paramètres de contexte
extraction de triplets C-R-C
Base de connaissances
JDM (Jeux de mots)
IA : interrogation de la base
mise en forme "DataViz"
Parcours d'apprentissage et champs conceptuels contextualisés
CJCPRAX 2019 – novembre 2019
12

Menu

A. État de l'art et objectifs de la recherche et substrats théoriques

B. Démarche, méthodes et premiers résultats

C. Participations et communications

D. Planning prévisionnel

La démarche

- Analyse de discours issus d'entretiens.
  - objectifs :

    Recueillir et questionner les pratiques (*Pélissier*) et les expériences des acteurs de la formation concernant la création d'un parcours de formation pour une discipline donnée.

Premiers résultats : l'analyse d'un processus "d'aide"

- Choix de privilégier une approche Système d'Information pour l'analyse du processus :

  L'aide à la transmission de connaissances est vue comme un processus interne au Système d'Information (processus de support au sens de Porter (1986))

  - le processus d'aide est décrit grâce à un formalisme de description et de documentation concernant :
    - les acteurs
    - les étapes
    - les produits
    - les indicateurs de qualité
  - on définit le processus de l'aide par son périmètre

Modèle de référence de travail

- Modéliser la connaissance dans la Théorie Anthropologique du Didactique (*Chaachoua 2010) :*

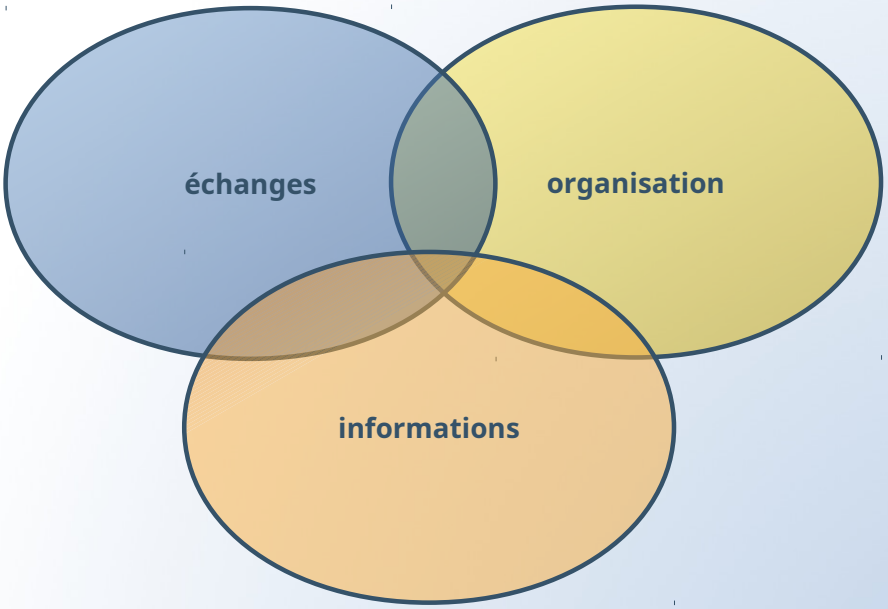
  - Toute activité humaine met en œuvre une organisation que Chevallard (1998) note [T / τ / θ / Θ] et qu'il nomme praxéologie, ou organisation praxéologique.
    - [T / τ] étant la pratique – ou encore le savoir- faire ;
    - [θ / Θ] le logos – ou encore le savoir.

→ On s'appuiera sur la praxéologie didactique avec comme type de tâche T les tâches d'étude propre à la discipline choisie.

Le processus

- C'est un ensemble d'activités reliées entre elles, réalisées par des acteurs et concourant à atteindre un objectif fixé. Il décrit l'enchaînement des activités nécessaires à la production des résultats.
- Qui fait quoi et comment :
  - Les acteurs qui interviennent ;
  - Les actes réalisés (transformation d'information) ;
  - Selon quel ordre, quelles séquences (mise en évidence des frontières entre acteurs) ?
- Comment le font-ils :
  - Description des actes incluant les outils de traitement

3 vues - 3 paradigmes



Vue "échanges"

C'est une vue extérieure au processus. Elle décrit les échanges d'information entre:

- les interlocuteurs à qui ce processus destine les résultats de son action.
- tous les acteurs dont il dépend pour produire ces résultats.

- C'est une vue axée sur les échanges

Vue "organisationnelle/fonctionnelle"

C'est une vue interne, elle décrit :

- Les acteurs (ou groupe d'acteurs) physiques ou organisationnels qui collaborent à l'atteinte des résultats (objectifs). Ces acteurs sont illustrés par leurs compétences (domaine de responsabilité, pouvoir de décision).

- Elle met en évidence la façon dont l'organisation prend en charge ses missions

Vue "informationnelle"

- L'ensemble des informations manipulées (les concepts et leurs inter-relations) qui constituent un micro-univers : la représentation restreinte du monde dont s'est dotée l'organisation pour accomplir ses missions.
- La démarche consiste à identifier, extraire, et modéliser l'information
- L'information n'est pas une information structurée : elle a donc besoin d'être modélisée, formalisée afin de simplifier son analyse.
- On cherche à en décrire son cycle de vie, depuis sa production jusqu'à sa délivrance.

Synthèse et premiers livrables

Le recueil et l'identification des informations récoltées dans le cadre de ces trois paradigmes (les trois vues) et articulées les unes par rapport aux autres feront l'objet d'un guide de référence du processus considéré :

→ Le "référentiel d'aide"

Choix des outils : les principes

Une chaîne de traitement numérique résolument basée sur des outils "open-source" et ouverts permettant des traitements multiples au travers

- d'outils dédiés (RQDA, Tropes, Iramuteq, TXM) ;
- de langages orientés vers le traitement de données et l'analyse statistique (R ou Python).

RQDA - fonctions "avancées"

- Un ensemble d'items accessibles au travers de menus contextuels relatifs :
  - aux fichiers importés
  - aux codes créés
  - aux catégories
  - aux mots apparaissant dans le corpus
  - aux typologies de fichier
- Toutes les variables créées peuvent être documentées et/ou journalisées

- Chaque concept/catégorie extrait est projeté dans le réseau lexical
- Objectifs :
  - Construire sur un réseau lexical généraliste un réseau lexical dédié à la didactique des sciences physiques (Thermodynamique) et d'aide à l'apprentissage.
  - Ce réseau combinera poids et annotations sur des relations typées entre des termes et des concepts. Il sera à l'origine de référentiels devant aboutir à la proposition de parcours de formation.

Description du corpus actuel

- Acteurs de l'enseignement supérieur :
  - enseignants et chercheurs (universités françaises et canadiennes)
  - ingénieurs pédagogiques (universités françaises, écoles et grands établissements)
  - étudiants (Québec)
- 21 entretiens
- modalités :
  - présentiel
  - téléphone
  - Skype

Guides d'entretien

- Deux guides d'entretien dynamiques permettant de s'adapter aux divers interlocuteurs avant et pendant les entretiens (cartes heuristiques)

- L'adaptation des guides se déroule de manière incrémentale selon les catégories émergentes d'une *analyse qualitative ancrée.(Glaser, al. 1968, Lejeune 2016)*

Guide de codage

- Il s'appuie sur les deux niveaux de catégorie proposés dans le cadre de RQDA
- une "condensation" des catégories d'analyse à partir des étiquettes assignées à des portions de corpus illustrant une typologie de l'aide réellement mise en œuvre.
- Un codage lié à la vue organisationnelle/fonctionnelle
  - identification des acteurs
  - relations entre acteurs
  - identification des actions
  - séquençage des actions
- un codage lié à la vue informationnelle permettant l'identification des "objets traités".

Première catégorisation *(provisoire)* d'aide

- Catégories d'aide identifiées :
  - Mise en confiance et empathie
  - Exploration : susciter une situation, émettre des hypothèses, faire préciser un mot ou un concept.
  - Nature de la construction à partir de la didactique de la discipline : poser et hiérarchiser les concepts, mettre à jour les contenus.
  - Définition du champ d'intervention dans le dispositif pédagogique proposé : identifier l'accompagnement dans ses objectifs identifier le paramètre institutionnel.
  - Relation entre l'apprenant, son apprentissage et la personne aidée : définir les prérequis, définir les objectifs, gérer les profils d'apprenant.
  - Appui à la pratique pédagogique : positionner une action dans une démarche d'apprentissage, qualifier et évaluer une pratique, relier un propos à une théorie de l'apprentissage.

Catégories d'aide (1/2)	Exemples d'expression
Mise en confiance et empathie	Aider le locuteur dans son expression : *"...C'est votre point de vue qui m'intéresse.."* Marquer la qualité de son écoute : *"OK," "Je vois", "Tout-à-fait"* Répéter un mot ou une expression en écho-miroir : *"...donc la didactique c'est essayer de structurer un enseignement ou tout au moins des connaissances pour pouvoir atteindre un objectif de formation...."*
Exploration	Susciter une situation ou une idée : *"...Ce qui serait intéressant effectivement, c'est qu'il existe un moyen de les inciter à faire ces liens et tout au moins leur montrer de tels liens ..."* Émettre des hypothèses : *"... et ça cela veut dire que les étudiants ne savent pas travailler.."* Faire préciser un mot ou un concept : *"...Quelle est votre définition de la didactique ?..."*
Nature de la construction à partir de la didactique de la discipline	Hiérarchiser les concepts : *"...Et la hiérarchisation des concepts à acquérir : il faut apprendre telle chose avant telle chose. etc..."* Mettre à jour des contenus : *"...comment vous introduisez les apports de la recherche dans votre enseignement ?..."* Poser les concepts : *"... pouvez-vous me définir votre définition à vous , celle qui vous semble la plus pertinente de la différence entre didactique et pédagogie ?..."*

Catégories d'aide (2/2)	Exemples d'expression
Définition du champ d'intervention dans le dispositif pédagogique proposé	Identifier l'accompagnement dans ses objectifs : *"...Ce que l'on attends d'un ingénieur pédagogique c'est ça ? C'est uniquement : j'arrive mon corpus , débrouille toi pour mettre les bons outils qui vont me permettre de faire passer ce corpus...."* Identifier le périmètre institutionnel de l'accompagnement proposé : *"... C'est de ça dont vous parliez en parlant de "cadre"...."*
Relation entre l'apprenant, son apprentissage et la personne aidée	Définir des prérequis : *"On devrait être capable de dire que tel concept on devrait avoir tel ou tel prérequis pour pouvoir le comprendre"* Définir un objectif de formation : *"Un expert du domaine qui est capable de dire que la notion d'enthalpie doit être apprise que parce que on a appris quelque part ce que c'était que l'entropie et l'entropie que parce qu'on a su quelque part ce que c'était que l'énergie et que sais-je encore ?"* Gérer les profils d'apprenant : *"Ça veut dire qu'il faut déjà faire une analyse de l'apprenant et du profil de l'apprenant avant la formation ?"*
Appui à la pratique pédagogique	Positionner une action dans une démarche d'apprentissage : *"Vous dites que les enseignants de manière général connaissent le fond, que vous n'intervenez pas sur le fond, cela dit comment adapter les outils au fond."* Qualifier une pratique : *"Donc le positionnement de l'ingénieur pédagogique c'est un accompagnateur de l'ensemble des acteurs de la formation: aussi bien aussi bien apprenant qu'enseignant"* Relier un propos à une théorie de l'apprentissage : *"tu t'appuies beaucoup sur les échanges et la découverte. Tu fais du socioconstructivisme"*

Ces catégories constituent des éléments clés pour :

- une base à l'alimentation du guide visant à faire expliciter par la personne interviewée la nature des activités qu'elle mène en tant qu'enseignant produisant seul un cours ou en tant qu'ingénieur pédagogique aidant un enseignant à structurer son cours.

- une mise en évidence d'éventuelles ruptures dans la chaîne de l'organisation des connaissances en vue de les transmettre.

- alimenter la démarche de production d'un outil informatique intégré destiné aux enseignants pour les aider à structurer les connaissances à transmettre et aux ingénieurs pédagogiques pour les aider dans leur travail d'aide à l'enseignant.

Bibliographie sommaire

- ARTIGUE, Michèle, 1989. *Épistémologie et didactique.* [en ligne]. S.l. : IREM. Cahier de DIDIREM, 3. Disponible à l'adresse : <http://docs.irem.univ-paris-diderot.fr/up/publications/IPS97066.pdf>.
- BALACHEFF, Nicolas et MARGOLINAS, Claire, 2003. cK Modèle de connaissances pour le calcul de situations didactiques. In : *XII° école d'été de didactique des mathématiques* [en ligne]. S.l. : La Pensée Sauvage éditions. 2003. p. 1–32. [Consulté le 20 juin 2017]. Disp
- BROUSSEAU G., 1998, Théorie des situations didactiques, 395p, ed. La Pensée Sauvage, Grenoble disponible à l'adresse : <https://hal.archives-ouvertes.fr/hal-01139408/>.
- CHAACHOUA, Hamid, 2010. La praxéologie comme modèle didactique pour la problématique EIAH. Etude de cas: la modélisation des connaissances des élèves. HDR. S.l. : Université de Grenoble.
- CHEVALLARD, Yves, 1991. La transposition didactique. S.l. : La pensée sauvage Grenoble. ISBN 2-85919-078-3.
- CHEVALLARD, Yves, 2003. Didactique et formation des enseignants. In : Journées d'études INRP-GÉDIAPS. S.l. : s.n. 20 mars 2003. p. 14.
- GIORDAN, André et DE VECCHI, Gérard, 1987. Les origines du savoir. In : *Neuchâtel, Delachaux et Niestlé*. 1987.
- GLASER, Barney G. and Strauss, Anselm L. and Strutzel, Elizabeth, The Discovery of Grounded Theory; Strategies for Qualitative Research, 1968
- LAFOURCADE, Mathieu et JOUBERT, Alain, 2008. Jeuxdemots: un prototype ludique pour l'émergence de relations entre termes. In : *JADT'08: Journées internationales d'Analyse statistiques des Données Textuelles* [en ligne]. S.l. : s.n. 2008. p. 657–666. [Consulté le 9 mars 2016]. Disponible à l'adresse : <https://hal.archives-ouvertes.fr/lirmm-00358848/>.
- LAFOURCADE, Mathieu et JOUBERT, Alain, 2018. Si beaucoup d'oiseaux volent, peut-on en déduire que tous les oiseaux volent ? Production endogène de règles déductives dans un réseau lexico-sémantique. In : *Proceedings TALIA 2018*. Nancy : s.n. 6 juillet 2018. p. 10.
- LEJEUNE, Christophe, Manuel d'analyse qualitative: analyser sans compter ni classer, 2016
- PELISSIER, Chrysta, Des Aides à l'apprentissage Pour La Formation à l'Université, 2011

Merci de votre attention...

... questions, précisions, suggestions ?

1.5 Annotation et exploitation numérique d'un corpus d'écrits scolaires corrigés en France et en Italie

Sara Mazziotti - (présentation)

Annotation et exploitation numérique d'un corpus d'écrits scolaires corrigés en France et en Italie

Sara Mazziotti

Université Sorbonne Nouvelle – Paris 3, ED 268, CLESTHIA, 4 rue des Irlandais Paris, France
sara.mazziotti@studio.unibo.it

Résumé. À partir d'un corpus de textes d'élèves de CE2 et de CM2 recueillis en France et en Italie, nous investiguerons les pratiques de correction des enseignants qui se confrontent avec deux systèmes linguistiques ayant un taux de correspondance entre graphème et phonème très différent. La numérisation de ces écrits, prenant en compte l'aspect génétique du texte, rend possible, d'une part, l'évaluation du taux de remaniement du brouillon de la part de l'élève et, donc, la prise en compte des différentes corrections apportées par l'enseignant lors de l'étape de réécriture. D'autre part, elle permet le repérage et la classification des modifications et des commentaires des enseignants apportés dans le brouillon. Sur quels aspects linguistiques en particulier se focalisent-ils ? Et quel est le taux de prise en compte des corrections par l'élève lors de l'étape de réécriture ?

Mots-clés. Ecrits scolaires, pratiques de correction, génétique textuelle, transcription, réécriture.

Abstract. Based on a corpus of texts from third and fifth graders collected in France and Italy, we will investigate the correction practices of teachers who are confronted with two linguistic systems with a very different correspondence rate between grapheme and phoneme. The digitization of these writings, taking into account the genetic aspect of the text, makes it possible, on the one hand, to evaluate the student's rate of reworking the draft and, therefore, to take into account the various corrections made by the teacher during the rewriting stage. On the other hand, it allows the identification and classification of changes and teacher comments made in the draft. Which linguistic aspects do they focus on? And what is the rate at which corrections are considered by the student during the rewriting stage?

Keywords. School writings, correction practices, textual genetics, transcription, rewriteing

1 Introduction

Cette étude est le résultat d'une série de questionnements de recherche que nous nous sommes posés pendant une collaboration avec l'équipe ECRISCOL, dirigée par Claire Doquet à l'Université Sorbonne Nouvelle – Paris 3. Le projet ECRISCOL a l'objectif de créer et de mettre à disposition des chercheurs et des enseignants une base de

données accessible en ligne constituée de textes d'élèves de l'école primaire au niveau universitaire. À chaque production écrite est associée une transcription en version TEI qui met en évidence tous les éléments supprimés, ajoutés, déplacés par l'élève ou par l'enseignant. À partir du même protocole de transcription, nous nous intéresserons dans notre thèse aux corrections des enseignants en particulier, afin de dessiner plusieurs « postures de correction »[1] de maitres et maitresses de CE2 et de CM2 en France et en Italie. Notre point de départ est le classement avancé par Jean-Luc Pilorgé qui distingue cinq postures de correction sur la base du type d'intervention de l'enseignant : le « gardien du code » s'intéresse principalement à la grammaire et à l'orthographe ; le « lecteur naïf » commente la « représentation du monde » de l'élève ; la posture de « stimulus réponse » vérifie le « respect de la tâche à accomplir » et le « réinvestissement de savoirs requis par le sujet » ; l'éditeur s'appuie sur le « déjà-écrit » et vise à « l'amélioration du texte » ; le critique commente le texte perçu comme une « construction, un objet de réflexion ». Dans la présentation des exploitations automatiques des transcriptions à l'aide du logiciel de textométrie MkALign[2] et de scripts informatiques conçus pour cette étude, nous nous focaliserons en particulier sur les postures de « gardien du code » et de « lecteur naïf ».

2 Hypothèse et méthodologie

Notre hypothèse est qu'un système linguistique avec une correspondance très faible entre graphème et phonème, comme dans le cas de la langue française, amène les enseignants à corriger plus particulièrement l'aspect orthographique du texte. En revanche, un système linguistique plus régulier, comme dans le cas de l'italien, permet aux enseignants de retrouver moins d'erreurs orthographiques dans les copies et, par conséquent, de proposer plus fréquemment des commentaires qui portent sur la cohérence d'un texte ou sur le contenu. Nous avons proposé dans huit classes de CE2 et de CM2 italiennes et françaises la consigne d'écriture : « Que feras-tu quand tu seras grand ? Raconte une de tes journées ». Les élèves devaient rédiger pendant une heure un premier texte d'une façon individuelle et, après une semaine environ, ils devaient relire le texte corrigé par leur enseignant pour en proposer une version définitive. Les deux versions ont été ensuite transcrites, exploitées automatiquement et relationnées aux réponses fournies par les enseignants à un questionnaire investiguant leurs pratiques de correction. Nous avons recueilli également une série de métadonnées concernant les élèves (langue maternelle, lieu de naissance, catégorie socio-professionnelle des parents, etc.), à travers une fiche que les parents devaient signer pour autoriser l'exploitation des copies anonymisées pour des fins de recherche.

[1] Jean-Luc Pilorgé, « Un lieu de tension entre posture de lecteur et posture de correcteur : les traces des enseignants de français sur les copies des élèves ». In : Pratiques, n° 145-146, 2010

[2] Serge Fleury, Paris 3

2.1 La transcription : un véritable outil d'analyse

La transcription, qui dans le cas de notre recherche est assez chronophage, puisqu'elle n'est pas automatique, constitue en réalité la première phase de notre analyse. Elle nous permet d'évaluer le nombre d'interventions de chaque enseignant et de repérer ses habitudes de correction qui seront ensuite vérifiées pendant l'analyse informatique. Pendant la transcription de la version définitive de l'élève, nous évaluons également le taux de variation du texte lors du passage à la réécriture : en général, un brouillon trop massivement corrigé amène l'élève à ne pas retravailler le contenu, mais à effectuer une mise au propre orthographique de la première version ou à l'abandonner pour en rédiger une différente. Dans la figure 1, nous proposons une mise en comparaison de deux extraits de copies de corrigées, tirées de notre corpus :

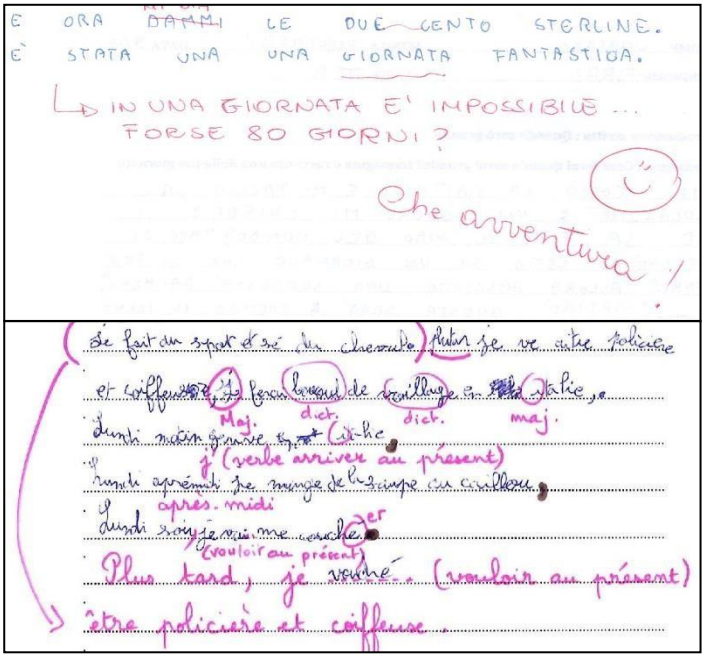


Fig.1 : En haut, un extrait de correction d'un enseignant italien de CE2 qui interpelle la représentation du monde de l'élève et qui adopte la posture de « lecteur naïf » pour commenter « En une seule journée il est impossible…peut-être en 80 jours ? ». L'élève avait raconté une série d'aventures et de voyages qu'il n'aurait pas pu effectuer en une seule journée. En revanche, en bas, un exemple de correction massive apportée par un enseignant français de CM2. Nous signalons la présence d'abréviations qui indiquent le type d'erreurs à l'élève (Maj : majuscule ; dict : dictionnaire...).

Il ne serait pas inutile de présenter alors brièvement les procédures d'écriture les plus fréquentes dans notre protocole, utilisé par le groupe de recherche ECRISCOL également, et que nous signalons avec les balises suivantes :

- <chevrons> si l'élève ou l'enseignant ajoute du contenu (ponctuation, lettre, mot, phrase,...) ;
- [crochets] si l'élève ou l'enseignant supprime du contenu ;
- [crochets]<chevrons> si l'élève ou l'enseignant supprime un premier élément pour le remplacer avec un nouvel élément (ex : [hier]<demain> si "hier" a été remplacé par "demain") ;
- {T2P#accolades} si l'enseignant ajoute un commentaire. "T2P" signale l'intervention de l'enseignant dans un temps successif au premier jet de l'élève et il apparaitra alors indépendamment de la procédure d'écriture qu'il applique ;
- {T2Pσ#accolades et sigma} si l'enseignant souligne un élément (ex : {T2Pσ#magazin} si l'enseignant souligne le mot pour indiquer qu'il y a une erreur d'orthographe.

3 L'analyse automatique des transcriptions

3.1 Le taux d'intervention des enseignants

Ne pouvant pas suivre les enseignants participant au projet pendant toute l'année scolaire, nous avons focalisé notre attention sur un devoir précis, notamment une consigne de réécriture, afin d'investiguer les différentes manières d'intervenir dans le texte d'élève qui devra être réélaboré individuellement. Comme nous l'avons précisé, toutes ces interventions sont signalées dans les transcriptions par le biais de balises qui peuvent donc être calculées automatiquement ou extraites et copiées dans un nouveau fichier. Le script que nous présenterons ici calcule le nombre d'interventions total de l'enseignant et le nombre exact de suppressions, ajouts, remplacements, soulignements et commentaires. Il fonctionne à partir d'un fichier txt qui rassemble l'intégralité des transcriptions des brouillons de ses élèves et il génère automatiquement un nouveau fichier avec les résultats de chaque catégorie (Fig.2). Nous avons ensuite calculé le nombre moyen d'interventions de chaque enseignant, en prenant en considérations le nombre de copies corrigées, afin de diviser de manière équilibrée les enseignants italiens et français en trois niveaux d'intervention : bas, moyen et élevé.

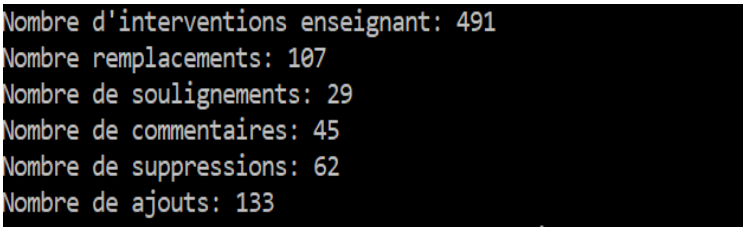


Fig.2 : Cet enseignant italien de CE2 a apporté 491 corrections totales sur 21 copies. Étant dans notre corpus 5,6 le nombre moyen par copie le plus bas (chez un enseignant italien) et 48,2 le plus élevé (chez un enseignant français), le degré d'intervention de cet enseignant est moyen.

Si la plupart des corrections sont locales, donc plus probablement associées à un mot bien précis, alors il est très probable que les enseignants avec un degré d'intervention élevé soient des « gardiens du code », en particulier, ceux qui n'apportent presque pas de commentaires. Nous analyserons successivement les différents types de commentaires chez les enseignants, puisque dans nos transcriptions les abréviations du type « orth, gr, temps », par exemple, et les commentaires sur le contenu sont transcrits de la même façon à l'aide des balises {T2P#accolades}.

3.2 La prise en compte des corrections chez l'élève

Afin d'évaluer le degré de réécriture du texte et d'étudier le type de modifications apportées par l'élève lors du passage du brouillon à la version finale, nous avons utilisé le logiciel MkAlign. Les élèves pouvaient en fait choisir de recopier le texte tel quel, le modifier en partie ou entièrement ou en rédiger un autre complètement différent de la première version. La mise en comparaison des deux textes nous permet de distinguer les éléments ajoutés, supprimés ou modifiés lors du passage à la version définitive et de relationner ces données aux postures de correction des enseignants (Fig. 3).

Fig. 3 : En vert, les éléments ajoutés ans la version définitive ; en rouge les éléments supprimés qui étaient présents dans le brouillon et en bleu les éléments déplacés ou modifiés.

Pour effectuer cette analyse, nous avons d'abord généré un script qui pouvait enlever automatiquement toutes les balises signalant les différentes procédures d'écriture et les corrections des enseignants. Le logiciel aurait sinon signalé en rouge presque tout le texte, puisque les corrections de l'enseignant étaient présentes uniquement dans le brouillon.

4 Conclusions et quelques résultats

À partir de notre classement des enseignants en trois niveaux selon leur degré d'intervention dans la copie, nous constatons que presque la totalité des enseignants français de CE2 et de CM2 applique un nombre élevé d'interventions ; en revanche, la plupart des enseignants italiens a un degré d'intervention bas. Cependant, le nombre

d'interventions locales est, dans les deux cas, plus élevé que les commentaires généraux en-tête, en marge ou en bas de page. Nous avançons alors l'hypothèse que la tendance à corriger surtout au niveau local ne dépende pas de la complexité du système linguistique : même les enseignants-correcteurs italiens se focalisent davantage sur l'aspect formel du texte. Cette volonté de signaler un maximum d'erreurs est-elle due à la perception de la tache de correction elle-même ? Et le degré d'intervention plus bas chez les enseignants italiens est-il motivé uniquement du fait qu'il y ait moins d'erreurs d'orthographe dans les copies ?

Une ultérieure analyse du type de commentaires proposés dans les deux contextes linguistiques nous permettra de répondre à cette dernière question : un nouveau script lancé à partir des transcriptions des brouillons corrigés génère un fichier dans lequel il propose une liste des commentaires de chaque enseignant. Portent-ils sur l'orthographe ou sur le contenu ? Quelle posture pourrions-nous alors attribuer à ces enseignants : toujours "gardien du code" ou "lecteur naïf » ?

Chez un enseignant italien de CE2, nous avons constaté que les deux postures peuvent co-exister au sein de la même copie et d'une façon assez équilibrée. En effet, il essayait de repérer un maximum d'erreurs d'orthographe, mais de proposer aussi en bas de page toujours un commentaire positif sur le contenu du texte, probablement pour communiquer à l'élève qu'il avait lu et apprécié ses idées (Fig. 1). Un « vrai lecteur » plutôt qu'un « lecteur naïf » dans ce cas-là : un « enseignant-lecteur » qui veut s'assurer que ses élèves aient compris qu'il a lu leurs textes.

De toute manière, cette duplicité de postures amène les élèves à ne pas abandonner leur premier texte et à le retravailler. Nous confirmons en effet dans notre étude aussi qu'un nombre très élevé de corrections inhibe la manipulation du texte de la part de l'élève qui passe à une mise au propre du brouillon ou à la rédaction d'un texte *ex novo*. Dans le cas de cet enseignant italien, l'équilibre entre ces deux postures permet aux élèves de rectifier les erreurs, mais au même temps de remanier le contenu. Notre objectif étant d'évaluer les pratiques de correction qui favorisaient davantage la réélaboration de la première version, dans le but d'aboutir à un texte de meilleure qualité.

Références

1. David, J., Guerrouache, C.,: Relire, analyser et réécrire ses écrits au début du primaire, Le français d'aujourd'hui, 29-49 (2018)
2. Doquet, C.: L'Écriture débutante. Pratiques scripturales à l'école élémentaire, Presses universitaires de Rennes, Rennes (2011)
3. Doquet, C., J. David, and S. F. : (Eds). Spécificités et contraintes des grands corpus de textes scolaires : problèmes de transcription, d'annotation et de traitement. In Corpus [Online], volume 16 (Special Issue). OpenEdition (2017)
4. Doquet, C., Enoiu, V., Fleury, S. et Mazziotti, S. : Problèmes posés par la transcription et l'annotation d'écrits d'élèves, Corpus [En ligne], 16 | 2017, URL : http://journals.openedition.org/corpus/2776

5. Fabre, C.: Les brouillons d'écoliers ou l'entrée dans l'écriture, Ceditel/L'atelier du texte, Grenoble (1990)
6. Fayol, M., Gombert, J.E., : Le retour de l'auteur sur son texte : Bilan provisoire des recherches psycholinguistiques », Repères, n° 73 (1987)
7. Pilorgé, J-L.: Un lieu de tension entre posture de lecteur et posture de correcteur : les traces des enseignants de français sur les copies des élèves. Thèse de doctorat sous la direction d'Annie Rouxel, Rennes, (2008)
8. Projet ECRISCOL: http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/
9. Logiciel MkAlign (Serge Fleury): http://www.tal.univ-paris3.fr/mkAlign/
10. Logiciel LeTrameur (Serge Fleury): http://www.tal.univ-paris3.fr/trameur/

Annotation et exploitation numérique d'un corpus d'écrits scolaires corrigés en France et en Italie

Sara Mazziotti

Encadrement : Université Paris 3 Sorbonne-Nouvelle, ED 268 (Pr. Claire Doquet)

Cotutelle : Université de Bologne – Cristiana De Santis

Colloque Jeunes Chercheurs PRAXILING 2019 :
Représentations et transmission des connaissances à la lumière de l'innovation numérique
7-8 nov. 2019
Montpellier (France)

UNIVERSITÉ SORBONNE NOUVELLE PARIS 3

Membre fondateur de Sorbonne Paris Cité

Sommaire

I. Le point de départ : EcriScol

II. Le projet de thèse

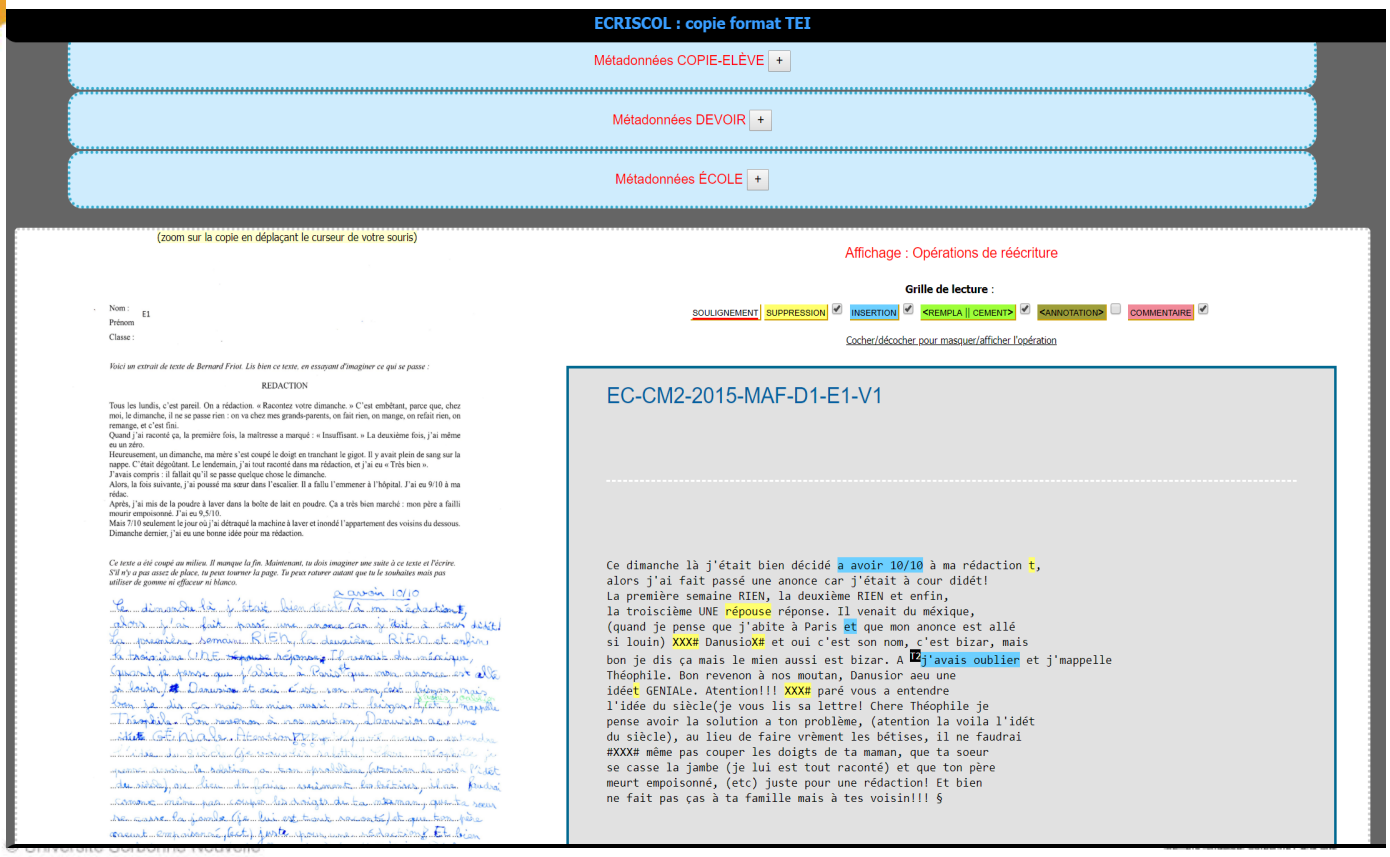
III. Le défi de transcription

IV. L'exploitation numérique

V. Premiers résultats

© Université Sorbonne Nouvelle

I. Le point de départ : EcriScol (Claire Doquet, Paris 3)



II. Le projet de thèse

Repérage de toutes les interventions des enseignants :

- Quelles sont les « postures de correction » (Pilorgé, 2010) les plus fréquentes ?
- De quelle façon le système linguistique influence-t-il les pratiques de correction des enseignants ?
- Comment les enseignants perçoivent-t-ils leur propre manière de corriger ?

Deux écoles françaises et deux écoles italiennes

Niveaux CE2 et CM2

LA CONSIGNE

Que feras-tu quand tu seras adulte?

DE REDACTION

Questionnaire enseignant, métadonnées de l'élève et du devoir

Analyse des données à l'aide des logiciels de textométrie **LeTrameur** et MkAlign

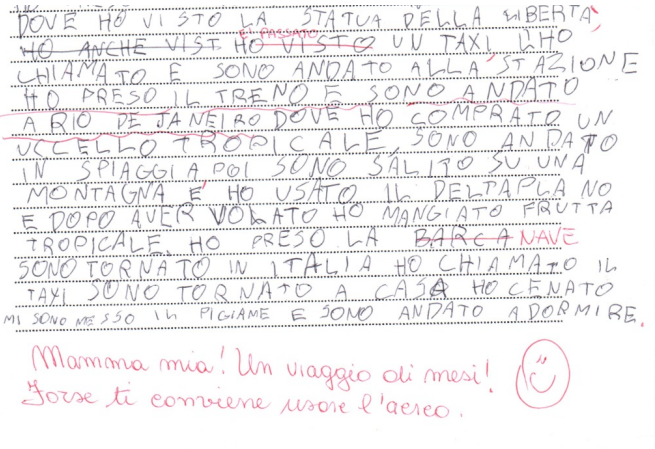
Transcription et annotation des productions des élèves

© Université Sorbonne Nouvelle

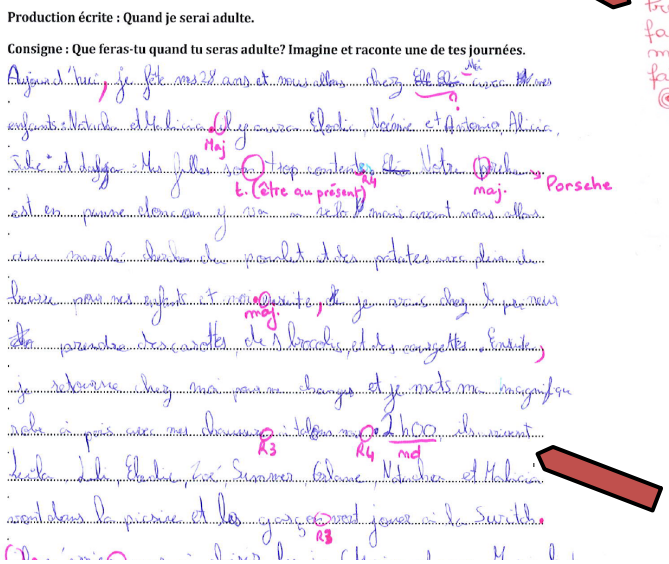
UNIVERSITÉ SORBONNE NOUVELLE PARIS 3

Membre fondateur Sorbonne Paris Cité

Lecteur-naïf
(Pilorgé, 2010)



EC-CM2-2017-FOR1-D1-E5-V1-002



Gardien du code
(Pilorgé, 2010)

EC-CM2-2017-JLF2-D1-E6-V1

© Université Sorbonne Nouvelle

III. Le défi de transcription

<était>_<étais> : correction orthographique

£ : renvoi involontaire, fin de page

[suppression]

//ajout//

T2# : action en temps 2

§ : fin de paragraphe, renvoi à la ligne volontaire

T2P# : action du prof (en temps 2)

{T2Pσ#e}: soulignements de l'enseignant

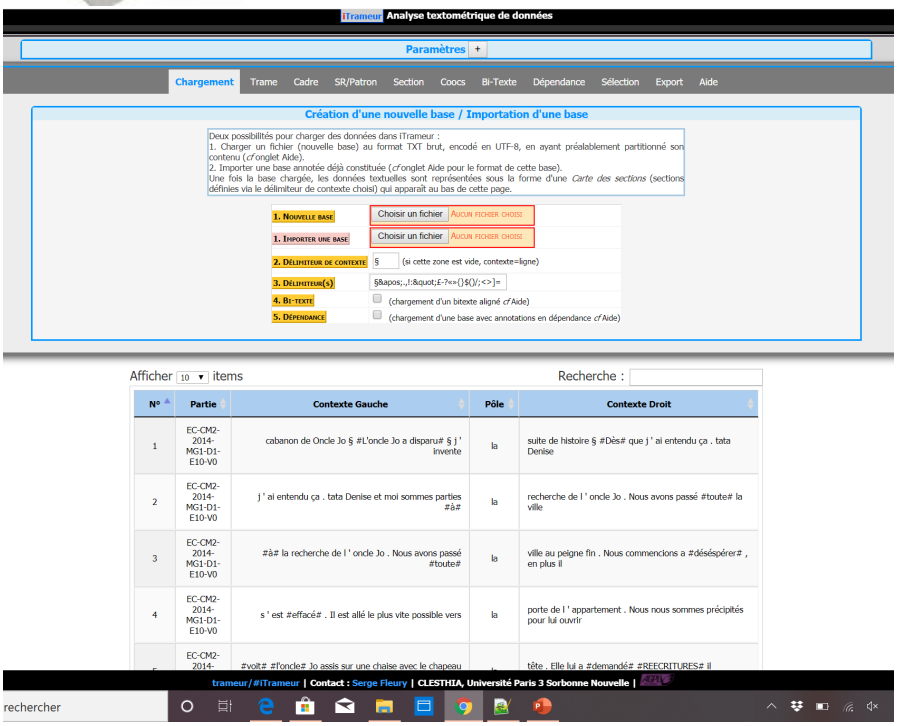
EC-CM2-2017-JLF2-D1-E6-V1

Aujourd'hui //T2P#,// je fête mes 28 ans et nous allons chez {T2Pσ#®[Elle]//Elé//® //Mai//}{T2P#?} avec [#X#]mes£enfants: Natacha et Malicia //T2P#.// <{T2Pσ#i}{T2P#maj}l>_<Il> y aura f_E, f_N et h_A, f_A.£
f_J et f_L. Mes filles <son{T2Pσ#}{T2P#t.(être au présent)}>_<sont> trop
<content{T2Pσ#}{T2P#R4}//T3#es//>_<contentes> [Eléa] Notre {T2Pσ#p}{T2P#maj.}orche //T2P#Porsche//£
est en panne donc on y va <a>_<à> vélo [#X#] mais avant nous allons£
au marché chercher du poulet et des patates avec plein de£
beurre pour mes enfants et moi //T2P#.// <{T2Pσ#e}{T2P#maj.}nsuite>_<Ensuite> //T2P#,// [de] je vais chez le
<prineur>_<primeur>£
[ch]prendre des carottes, des brocolis, et des courgettes. Ensuite //T2P#,//£
je retourne chez moi pour me changer et je mets ma <magnifiqu>_<magnifique>£
robe à pois avec mes <chaussur{T2P#R3}>_<chaussures> à talons <noir{T2Pσ#}{T2P#R4}>_<noirs>.
{T2Pσ#2h00}{T2P#md}, ils arrivent£f_L, f_L, f_E, f_Z, f_S, f_G, f_N et f_M£vont dans la piscine et les
<garçon{T2Pσ#}{T2P#R3}>_<garçons> vont jouer à la Switch
//T2P#.//£<{T2Pσ#i}{T2P#maj.}l//T2P#s//>_<Ils> n'<arrive{T2Pσ#}{T2P#t.(accord)}>_<arrivent> pas à choisir
les <jeu>_<jeux> (Marioodyssey, Mario kart,£
Just, #XXX##, etc). Et nous on boit un café tranquillement en parlant£
de <tou{T2Pσ#s}>_<tout> est n'importe quoi. Ensuite les enfants <échang{T2Pσ#er}{T2P#t.
(accord)}>_<échangeaient> de <{T2Pσ#robe}{T2P#?}>_<robes>£ jusqu'<a>_<à> 8h§

© Université Sorbonne Nouvelle

IV. L'exploitation numérique

1. iTrameur (Fleury, Paris 3) :



- Concordance
- Calcul vocabulaires spécifiques
- Erreurs fréquentes
- Recherches par catégories grammaticales

© Université Sorbonne Nouvelle

UNIVERSITÉ SORBONNE NOUVELLE PARIS 3
Membre fondateur Sorbonne Paris Cité

2. Scripts informatiques

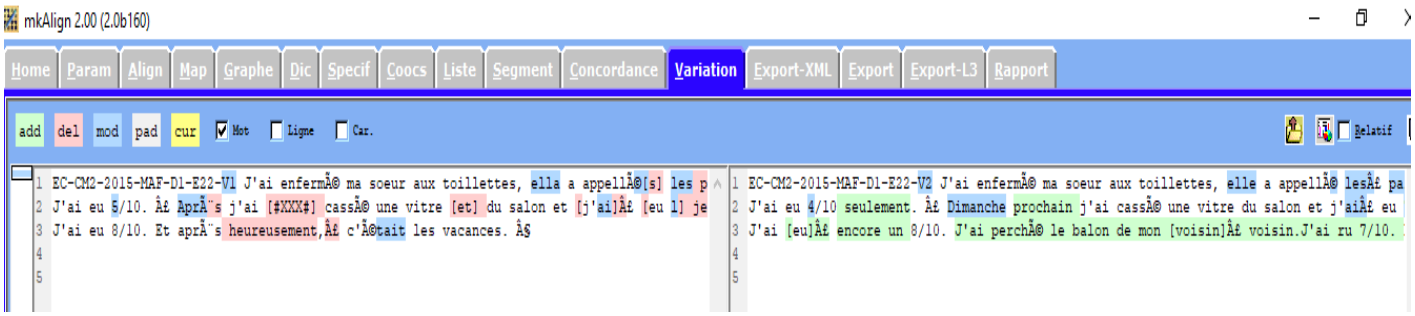
- Conversion en fichiers exploitables par iTrameur et incluant Treetagger ;
- Calcul des interventions de chaque enseignant qui seront mises en relation avec le nombre de copies, afin d'établir un nombre moyen d'interventions par copie :

```
Nombre d'interventions enseignant: 491
Nombre remplacements: 107
Nombre de soulignements: 29
Nombre de commentaires: 45
Nombre de suppressions: 62
Nombre de ajouts: 133
```

- Extraction de toutes les erreurs d'orthographe, des erreurs corrigées et des erreurs non corrigées par l'enseignant.

© Université Sorbonne Nouvelle

3. MkAlign (Fleury, Paris 3) :



Première version Version définitive

Hypothèse : les élèves italiens sont moins inhibés lors de la réécriture, ils remanient plus facilement le contenu ⥗ quelle place accorder au système linguistique et à la considération des erreurs d'orthographe ?

© Université Sorbonne Nouvelle

V. Premiers résultats

A. Nombre inférieur de corrections chez les enseignants italiens et qui portent plus souvent sur le **contenu**, sur l'organisation du texte ;

B. Présence toutefois de la posture du « **gardien du code** » chez les enseignants italiens également (vers un registre plus soutenu) ⇒ Il est beaucoup plus simple de corriger l'orthographe et la grammaire plutôt que de fournir des conseils pour améliorer la cohérence du texte, par exemple;

C. Entre 5,6 et 48,2 interventions par copie :

- Enseignants français : **entre 30,2 et 48,2 (niveau élevé) ;**
- Enseignants italiens : **entre 5,6 et 43 (niveau moyen et bas).**

© Université Sorbonne Nouvelle

Références bibliographiques

DOQUET, Claire (2003). *Etude génétique de l'écriture sur traitement de texte d'élèves de Cours Moyen 2*, Linguistics. Université de la Sorbonne nouvelle - Paris III.
DOQUET, Claire, ENOIU, Vanda, FLEURY, Serge, MAZZIOTTI, Sara (2017). Problèmes posés par la transcription et l'annotation d'écrits d'élèves, *Corpus 16*, 133-156.
FABRE, Claudine (2002). *Réécrire à l'école et au collège. De l'analyse des brouillons à l'écriture accompagnée*, ESF Editeur, 2002
PENLOUP, Marie-Claude (1994). *La Rature n'est pas un raté. Plaidoyer pour le brouillon*, Rouen, MAFPEN.
PILORGE, Jean-Luc (2010). Un lieu de tension entre posture de lecteur et posture de correcteur : les traces des enseignants de français sur les copies des élèves. *Pratiques* 145-146, pp. 85-103.

Site iTrameur : http://www.tal.univ-paris3.fr/trameur/iTrameur/
Site EcriScol : http://syled.univ-paris3.fr/ecriscol/CORPUS-TEST/

© Université Sorbonne Nouvelle

1.6 Vers une résolution des relations anaphoriques dans la communication électronique médiée

Hani Guenoune, Cédric Lopez, Guillaume Tisserant, Mathieu Lafourcade, Melissa Mekaoui - (présentation)

Vers une résolution des relations anaphoriques dans la communication électronique médiée

Hani Guenoune[1,2], Cédric Lopez[2], Guillaume Tisserant[2], Mathieu Lafourcade[1], and Melissa Mekaoui[2]

[1] LIRMM / Campus St Priest, 161 Rue Ada, 34090 Montpellier, France

[2] Emvista / Rond-Point Benjamin Franklin, 34000 Montpellier, France

Résumé La tâche de résolution des coréférences consiste à identifier toutes les unités linguistiques faisant référence aux mêmes entités, propriétés ou situations du monde réel. Dans la démarche d'acquérir de la connaissance à partir de données textuelles non structurées, il s'agit d'un maillon essentiel aux chaînes d'analyse automatique du langage naturel. L'efficacité des systèmes de résolution automatique dépend directement des ressources utilisées et du caractère rédactionnel du texte auquel ils s'appliquent. Après un sommaire des travaux existants, cet article présentera les enjeux d'une résolution automatique des anaphores dans un contexte de communication électronique en introduisant les verrous susceptibles d'être rencontrés et les pistes envisageables pour les résoudre.

Mots-clé: Anaphores · Coréférences · Anaphores abstraites

Abstract. The task of coreference resolution applied to written texts consists in finding all lexical parts that refer to the same real world entities, properties or situations. In the aim of extracting knowledge from unstructured textual data, this task plays an essential role within the typical natural language processing pipeline. The efficiency of automated resolution systems depends on the one part, on the ressources it relies on, and the redactional nature of the text on the other. After an overview of existing works, this paper will present the issues of dealing with anaphora in the context of electronic communication, the problems one can encounter and consider ways of working around them.

Keywords: Coreference · Anaphora · Abstract anaphora.

Introduction

L'anaphore, phénomène linguistique largement étudié, est approchée selon des angles différents. L'anaphore se définit en rhétorique comme la figure de style consistant à multiplier l'usage d'un syntagme dans le but de renforcer une affirmation ou encore rythmer des propos autour d'une notion centrale [54]. L'anaphore en grammaire est vue comme l'opération imposant une contrainte de dépendance d'interprétation à deux mentions distinctes, dont une première serait

indispensable à la compréhension de la seconde. Elle consiste en une consigne implicite qui dispense de l'interprétation sémantique d'un terme, suggérant plutôt, -pour son interprétation-, une entité lexicale distante. L'identification des référents d'un expression anaphorique est essentielle à la compréhension intégrale d'un discours, permettant notamment à l'allocutaire de connecter les informations relatives aux mêmes entités, la synthèse des caractéristiques citées à différents segments du discours mènent à une définition plus complète des entités dont traite l'énoncé.

Dans la mouvance actuelle de l'intelligence artificielle et du traitement automatique du langage naturel (TALN), plusieurs travaux se consacrent à la question d'identifier les références au sein d'un discours oral ou écrit, à repérer les entités qui sont notées à plus d'un endroit, référencées par plusieurs unités lexicales. L'idée étant de décider si une mention est introduite à l'auditeur, ou est la réapparition d'une notion, propriété ou situation préalablement définie.
La tâche de localisation des référents des syntagmes d'un discours occupe une position centrale dans l'architecture des systèmes technologiques visant à tirer profit du langage écrit ou oral. Son accomplissement est nécessaire aux outils de résumé et de traduction automatiques [56; 64], à l'analyse de sentiments [5], ou encore aux systèmes de questions-réponse [7].

La résolution automatique des anaphores se fait aujourd'hui sur la base des rôles grammaticaux des termes, leurs dépendances ainsi que leurs natures sémantiques. Ces exigences renforcent la difficulté de résolution de traitement dans les textes issus de la communication électronique. Dans ce même contexte, la variété des types d'anaphores observées conduit notamment à s'intéresser aux formes d'anaphores les plus répandues dans ce contexte. Nous notons par exemple une utilisation fréquente de l'anaphore abstraite qui consiste à employer des syntagmes (*shell nouns* [28]) qui renvoient vers des entités abstraites représentées syntaxiquement par une, ou un ensemble de phrases.

Cet article commencera par dresser un aperçu des méthodologies employées dans les travaux abordant la résolution automatique des anaphores ainsi que des jeux de données disponibles (*cf.* section 1). Puis, l'article mettra en avant les spécificités de cette tâche en contexte de communication électronique (*cf.* section 2).

1 Travaux antérieurs

1.1 Terminologie

Dans l'intention de capturer les spécificités des différents schémas de reprise, différentes appellations leur sont assignées. Une distinction émanant de [20] est réalisée par la paire de mots **exophore** et **endophore** dont les préfixes séparent respectivement les caractères externe et interne du contexte dans lequel se définit le référent.
Ainsi, les *endophores* établissent un renvoi vers des éléments internes au discours,

tandis que les *exophores* désignent des entités qui se définissent dans un contexte situationnel, extérieur à l'énonciation et au discours.
Anaphore et **cataphore** se distinguent par la localisation relative du terme « source », porteur de la sémantique qui se veut rapportée ou propagée.
L'anaphore admet un terme source *antécédent*, mentionné préalablement dans le discours avant d'être repris. À l'opposée, la cataphore attend l'introduction d'un terme source *conséquent*, apparaissant *a posteriori* de la mention de reprise.
Ce dispositif réussit une simplification des appellations, au moyen de l'acception d'une terminologie à deux classes distinctes (exophore/endophore), éliminant dès lors la confusion engendrée par la multiplication de nominations. Le terme *exophore* écarte les variantes équivalentes *déixis* et *référence*, ce qui permet au dernier de maintenir son usage élémentaire et générique, celui d'une opération fondamentale sur laquelle on pourrait asseoir la construction linguistique plus complexe de reprise.
La **référence** représentera ainsi une opération de *désignation* agissant entre l'état lexical d'un terme et un objet concret ou abstrait du monde du discours. Le **renvoi**, en revanche, admet un couple d'opérandes lexicales. Le terme renvoyé est soumis à une interprétation sémantique dépendant du contexte lexical interne.

1.2 Typologie.

Bien que l'anaphore se présente dans des formes très variées, les travaux portant sur la résolution de l'anaphore adoptent souvent des typologies élémentaires. Le choix des différentes catégories considérées est imposé d'abord par la langue étudiée, ensuite par les besoins des applications pour lesquelles le système est développé. On distingue également l'anaphore *interphrastique*, dans laquelle les termes liés sont cités dans deux phrases différentes, de l'anaphore *intraphrastique* ou l'antécédent est localisé au sein de la même phrase que le terme anaphorique. Les types le plus souvent considérés en TALN sont les suivants [40] :

— *L'anaphore pronominale* représente le type le plus répandu, et consiste en une reprise par un pronom anaphorique.
— *L'anaphore nominale ou lexicale* désigne le renvoi opéré par un nom commun à un antécédent pouvant être un synonyme ou de sémantique proche.
— *L'anaphore d'instance* représente le cas où l'expression anaphorique n'entretient pas de relation d'identité avec l'antécédent mais désigne une instance de la classe sémantique à laquelle appartient l'antécédent.

1.3 Ressources textuelles

L'élaboration de corpus de textes annotés en informations référentielles est nécessaire aux recherches sur la résolution des coréférences. La disponibilité de ces données permet l'apprentissage des descripteurs et des marqueurs qui induisent l'existence de liens anaphoriques. Elle rend également possible l'évaluation des systèmes automatiques de résolution, en testant leur efficacité sur une grande quantité de données.

Au delà de leur finalité applicative, ces collections de données sont précieuses pour les études descriptives du langage naturel. Il s'avère toutefois que la création de ces ressources n'est pas une tâche triviale, de multiples schémas d'annotation ont été proposés sans que la communauté ne puisse converger vers une solution unifiée.

Les corpus *MUC* et *ACE* [21; 13] ont été spécifiquement conçus pour les tâches relatives à l'extraction d'information. Les choix d'annotation ont été pensés de sorte à répondre aux besoins des applications d'extraction d'informations, parfois au détriment des définitions formelles de linguistique. Les modifieurs des syntagmes nominaux n'étaient pas pris en compte, seule la tête d'une phrase nominale était annotée, et les phrases verbales ne figuraient pas au sein des mentions annotées. *ACE* limitait également les types sémantiques des syntagmes nominaux considérés à sept types, pertinents au domaine d'extraction de relations [21].

Les corpus *MATE*, *GNOME* [47], *ARRAU* [48], ainsi que *OntoNotes* [51] visent à couvrir l'étude de la coréférence des points de vues théoriques et pratiques, ils ne se consacrent pas spécifiquement à une tâche en particulier et adoptent des schémas d'annotation qui prennent en compte une typologie des anaphores plus complète. Ils considèrent notamment un ensemble plus large de types grammaticaux des mentions à annoter. *MATE* se penche également sur les phénomènes d'ellipses et propose des annotations pour des relations anaphoriques non identitaires (anaphores possessives, associatives).

ANCOR est un corpus français annoté en informations coréférentielles, il consiste en une transcription de données orales. Ses grandes collections de données visent à représenter une variété de langages de domaines différents. Il est constitué de trois corpus distincts, les corpus *Accueil_UBS*, *OTG* et *ESLO*. Le dernier étant le plus grand des trois, il transcrit des échanges dans le cadre d'interviews socio-linguistiques. [43].

DEMOCRAT est la ressource la plus récente, annotée en informations référentielles, son apport consiste à prendre en compte les chaînes de références dans le processus d'annotation [30]. La sélection des textes s'est faite de manière à étudier puis capturer les variations des chaînes de référence en fonction des genres discursifs et des époques. Les textes contenus dans le corpus varient selon plusieurs critères ; époque, type de texte (narratif ou non-narratif), genre textuel.

1.4 Systèmes de résolution automatique

Un certain nombre de travaux donnent un large sommaire des approches automatiques à la résolution des expressions référentielles [59; 49; 45; 44]. Sans prétendre à une quelconque exhaustivité, cette section proposera une vue d'ensemble des recherches qui se sont penchées sur la question du traitement automatique des anaphores.

Méthodes à base de règles. Après l'exclusion des mentions non-anaphoriques [62; 10] et étant donné un pronom à résoudre, ces méthodes utilisent des filtres heuristiques dans le but d'écarter les antécédents incompatibles, tandis que les contraintes de préférence calculent un score aux candidats restants afin d'en sélectionner le meilleur.
Les préférences se basent sur des informations de différents niveaux linguistiques, avec un accent sur le niveau morpho-syntaxique [22; 6] et sur la théorie du centrage [18; 25; 61]. Quelques travaux définissent également des règles sémantiques en exploitant des ressources externes [17; 50].

Il s'est manifesté dans cette lignée de travaux symboliques, une tendance à remplacer les systèmes à base de connaissances (knowledge-rich) par des systèmes avec peu de connaissances ou de pré-traitements requis, qui cherchaient à se passer de toute sémantique ou de connaissance de sens commun [29], voire se passer complètement d'analyse morpho-syntaxique préalable [26; 3; 38].

Au delà de la récente disponibilité de grands volumes de données, la difficulté d'une pondération manuelle d'un système avec un grand nombre d'attributs incite au passage aux modèles d'apprentissage qui permettent une gestion plus efficace d'ensembles de descripteurs très grands. Dans [37], les auteurs rapportent que leur modèle d'apprentissage surpassait le meilleur paramétrage de leur modèle à base de règles. [69] relèvent toutefois le manque d'adaptabilité des systèmes statistiques aux différentes formes d'anaphores.
La performance des méthodes statistiques a poussé les systèmes à base de règles à repenser la manière dont l'ensemble des attributs était géré. [19] ordonnent les règles définies, qu'ils appliquent successivement en commençant par les plus importantes. L'idée de définir des séquences de filtres et de règles a été développée dans [52] puis dans [31]. Plus récemment, [32] proposent une approche hybride combinant des classifieurs statistiques à des couches symboliques.

Méthodes statistiques. Dans la fin des années 90, le domaine de la résolution d'anaphore en TALN a été marqué par une transition vers les systèmes d'apprentissage automatique. Les premiers modèles implémentaient des algorithmes génétiques [41; 39], ou en encore des réseaux bayésiens [16]. La performance des modèles statistiques s'est graduellement améliorée et finit par surpasser les résultats atteints jusque-là par les systèmes symboliques.

Les modèles mention-pair basés sur les mentions procèdent de droite à gauche en comparant la mention à résoudre avec les mentions précédentes, une à une, donnant une prédiction binaire sur l'anaphoricité de la paire.
Les stratégies utilisées variaient entre celles dites *link-first* [55; 58] et *best-first* [46; 67; 4].
La démarche consistant à vérifier les mentions par paires peut générer des chaînes violant la contrainte de transitivité de la relation de coréférence, ce qui aboutit dans un contexte global à des incohérences dans le regroupement des entités.

Les modèles entity-based basés sur les entités, profitent des vérifications opérées avec les mentions précédentes [35]. Celles-ci se révèlent intéressantes quand une simple paire de mentions ne suffit pas à se prononcer sur l'égalité de leurs références.

Les modèles de classement ranking models répondent à la limitation de la classification binaire qui ne donne pas d'intuition sur la raison du choix d'un antécédent au lieu d'un autre. Cette lacune est adressée en instaurant une compétition qui aboutit à un classement des candidats. [14; 68; 23; 11]

Les modèles d'apprentissage profond ont permis d'améliorer les systèmes de classement notamment en ce qui concerne la détection des mentions anaphoriques [33; 8; 9; 65; 66].

2 Coréférences dans la communication électronique

La communication électronique prend diverses formes, plus ou moins formelles, variant des messages courts (SMS et plateformes de communication instantanée) aux e-mails professionnels. Ces formes de communication obéissent chacune à un ensemble de règles rédactionnelles nouvelles, qui diffèrent des standards littéraires traditionnels. Dans cette partie, nous aborderons les aspects théoriques et techniques appliqués à la communication électronique.

2.1 Aspects théoriques

Les formes de reprise observées dans le langage diffèrent dans une énonciation orale ou écrite et sont imposées par les standards de rédaction du domaine traité. La communication électronique adopte un style qui se rapproche du langage oral, favorisant les ellipses et mettant en avant des types d'anaphores qui peuvent se révéler difficiles à traiter. Nous abordons dans la suite deux phénomènes fréquents dans la communication électronique : les anaphores synthétiques et abstraites ainsi que les ellipses.

Anaphores synthétiques et abstraites. Elles font l'objet de diverses études linguistiques qui démontrent la particularité des propriétés des types sémantiques de leurs antécédents, le degré d'abstraction et les propriétés générales des discours dans lesquels elles s'utilisent [2; 63].
Comme dans (1), l'anaphore abstraite est un renvoi à un objet sémantique abstrait dont la représentation syntaxique est une phrase nominale ou verbale [2]. Parmi ces objets sémantiques peuvent figurer les événements, les situations, les propositions, ou encore les faits [1].

> *Il y a des problèmes concernant le développement de l'application.*
> *Cette difficulté doit être surmontée rapidement.* (1)

Compte tenu de la difficulté de la tâche d'annotation d'anaphores abstraites, peu de corpus sont disponibles [12]. Les études les plus anciennes de résolution automatiques se sont spécialisées dans le langage oral [15; 57; 42]. Récemment, [24; 34; 53] étudient la résolution des coréférences des évènements, et les anaphores abstraites de manière générale [36; 27; 28].

Par opposition à l'anaphore (pro)nominale, la résolution de l'anaphore abstraite s'avère plus contraignante dans le sens où un certain nombre de marqueurs permettant l'extraction et la sélection des bons candidats deviennent inutilisables. Les attributs d'accord de genre et de nombre entre l'anaphore et l'antécédent ainsi que leur correspondance lexicale fournissent des pointeurs très avantageux dans la discrimination entre les antécédents potentiels. Ces vérifications n'étant évidemment pas applicables à une énonciation phrastique, l'absence de ces indicateurs constitue un inconvénient majeur dans l'ambition d'identifier et de résoudre un lien anaphorique.

Ellipses. L'emprunt au style oral conduit à une utilisation fréquente d'ellipses de différents types. On retrouve des ellipses nominales (2) notamment dans l'apposition de propositions contrastives décrivant différentes caractéristiques d'une même notion. Des ellipses du sujet et du verbe sont également récurremment observées (resp (3), (4)).

Il y a de petites bouteilles, tu veux que je prenne des grandes ? (2)

Mathieu passe au dépot et revient ce soir (3)

Mathieu arrivera en voiture et Melissa à pied (4)

L'équipe est satisfaite de la v2, nous sommes tous soulagés. (5)

Ton bureau est plus grand que le mien. (6)

Il convient également d'évoquer l'utilisation des *anaphores associatives* (5) qui s'emploient entre des termes qui dépendent l'un de l'autre mais qui ne sont pas coréférents (liés par la relation sémantique de méronymie par exemple). Ou encore les *anaphores d'appartenance (dissociées)* (6) dans lesquelles le terme anaphorique désigne une instance différente de la classe sémantique à laquelle appartient l'antécédent.

2.2 Difficultés techniques

Dans cette section, nous listons les points difficiles que nous avons identifiés dans le cadre de la résolution automatique d'anaphores dans la communication électronique.

Écrits non standards. Dans la communication électronique, les écrits sont qualifiés de "non standards" ; plusieurs typologies ont été dressées pour classer les phénomènes linguistiques [60]. La fiabilité d'un certain nombre de descripteurs linguistiques est donc diminuée. La mise en place d'un système de normalisation et de correction orthographique peut être bénéfique à la résolution des anaphores.

Références. Les communicants peuvent échanger des objets numériques (fichiers textes, vidéos, images, *etc.*). La résolution des références doit analyser le contenu de ces fichiers et/ou de leurs méta-données.

Support de communication. Un échange textuel entre deux ou plusieurs communicants est partitionné en plusieurs segments de textes formant un fil de discussion. La portée des renvois est étendue au-delà du simple corps du texte. L'anaphore doit dans ce cas être résolue au sein d'une collection de textes qui n'ont pas de liens apparents.

Absence potentielle d'antécédents. L'antécédent d'une mention anaphorique peut être cité dans le texte même, dans le fil de discussion, ou encore dans un échange verbal entre les participants. La présence d'exophores, et l'appui sur le contenu de conversations orales sont des éléments qui contribuent à la complexité de la tâche et nécessitent de repenser les structures des ressources utilisées.

Absence de corpus annotés. Nous avons mis en évidence en section 1.3 que, malgré les ressources émergentes visant à contribuer à la résolution automatique des anaphores, aucun jeu de données dans un contexte de communication électronique n'est disponible.

Conclusion

La résolution automatique des anaphores est une tâche dont la difficulté varie selon les types des anaphores étudiés et le caractère rédactionnel des textes auxquels elle s'applique.
Étant donnée la particularité de la structure des échanges électroniques, il est important de construire un corpus annoté en informations référentielles. Un tel jeu de données serait un point de départ à la résolution des anaphores dans la communication électronique :

1. Sa conception nécessitera une réflexion sur la modélisation de la structure des échanges.
2. Le corpus annoté permettra d'identifier les types d'anaphores les plus fréquents dans le contexte de la communication électronique.
3. Le corpus pourra servir de jeu d'apprentissage, de développement et d'évaluation

Une attention particulière doit être accordée aux anaphores abstraites à antécédents phrastiques qui doivent faire l'objet d'études descriptives plus détaillées. Enfin, des efforts de normalisation des textes doivent également être fournis afin de corriger les erreurs grammaticales et lexicales et permettre une analyse morpho-syntaxique satisfaisante.

Références

[1] Amsili, P., Denis, P., Roussarie, L., and Umr, C. P. (2005). Anaphores abstraites en français : représentation formelle.

[2] Asher, N. (1993). Reference to abstract objects in discourse. In *Studies in linguistics and philosophy*.
[3] Baldwin, B. (1997). Cogniac : High precision coreference with limited knowledge and linguistic resources.
[4] Bengtson, E. and Roth, D. (2008). Understanding the value of features for coreference resolution. In *EMNLP*.
[5] Cambria, E. (2016). Affective computing and sentiment analysis. *IEEE Intelligent Systems*, 31 :102–107.
[6] Carbonell, J. G. and Brown, R. D. (1988). Anaphora resolution : a multi-strategy approach. In *COLING 1988*.
[7] Castagnola, L. (2002). Anaphora resolution for question answering.
[8] Clark, K. and Manning, C. D. (2016a). Deep reinforcement learning for mention-ranking coreference models. In *EMNLP*.
[9] Clark, K. and Manning, C. D. (2016b). Improving coreference resolution by learning entity-level distributed representations. *ArXiv*, abs/1606.01323.
[10] Danlos, L. (2006). Ilimp : Outil pour repérer les occurences du pronom impersonnel il. In *Actes de TALN 2006, Louvain, Belgique*.
[11] Denis, P. and Baldridge, J. (2008). Specialized models and ranking for coreference resolution. In *EMNLP*.
[12] Dipper, S. and Zinsmeister, H. (2012). Annotating abstract anaphora. *Language Resources and Evaluation*, 46 :37–52.
[13] Doddington, G. R., Mitchell, A., Przybocki, M. A., Ramshaw, L. A., Strassel, S., and Weischedel, R. M. (2004). The automatic content extraction (ace) program - tasks, data, and evaluation. In *LREC*.
[14] Durrett, G. and Klein, D. (2013). Easy victories and uphill battles in coreference resolution. In *EMNLP*.
[15] Eckert, M. and Strube, M. (2000). Dialogue acts, synchronizing units, and anaphora resolution. *J. Semantics*, 17 :51–89.
[16] Ge, N., Hale, J., and Charniak, E. (1998). A statistical approach to anaphora resolution. In *VLC@COLING/ACL*.
[17] Godbert, E. and Benoit, F. (2017). Détection de coréférences de bout en bout en français.
[18] Grosz, B. J., Joshi, A. K., and Weinstein, S. (1995). Centering : A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21 :203–225.
[19] Haghighi, A. and Klein, D. (2009). Simple coreference resolution with rich syntactic and semantic features. In *EMNLP*.
[20] Halliday, M. A. K. and Hasan, R. (1976). Cohesion in english.
[21] Hirschman, L. and Chinchor, N. (1997). Coreference task definition (v3.0, 13 jul 97).
[22] Hobbs, J. R. (1978). Resolving pronoun references.
[23] Iida, R., Inui, K., Takamura, H., and Matsumoto, Y. (2003). Incorporating contextual cues in trainable models for coreference resolution.
[24] Jauhar, S. K., Guerra, R., Pellicer, E. G., and Recasens, M. (2015). Resolving discourse-deictic pronouns : A two-stage approach to do it. In **SEM@NAACL-HLT*.

[25] Kameyama, M. (1997). Intrasentential centering : A case study. *ArXiv*, cmp-lg/9707005.
[26] Kennedy, C. and Boguraev, B. (1996). Anaphora for everyone : Pronominal anaphora resolution without a parser. In *COLING*.
[27] Kolhatkar, V. and Hirst, G. (2014). Resolving shell nouns. In *EMNLP*.
[28] Kolhatkar, V., Zinsmeister, H., and Hirst, G. (2013). Interpreting anaphoric shell nouns using antecedents of cataphoric shell nouns as training data. In *EMNLP*.
[29] Lappin, S. and Leass, H. J. (1994). An algorithm for pronominal anaphora resolution. *Computational Linguistics*, 20 :535–561.
[30] Lattice, LiLPa, ICAR, and IHRIM (2019). Democrat. ORTOLANG (Open Resources and TOols for LANGuage) –www.ortolang.fr.
[31] Lee, H., Chang, A. X., Peirsman, Y., Chambers, N., Surdeanu, M., and Jurafsky, D. (2013). Deterministic coreference resolution based on entity-centric, precision-ranked rules. *Computational Linguistics*, 39 :885–916.
[32] Lee, H., Surdeanu, M., and Jurafsky, D. (2017a). A scaffolding approach to coreference resolution integrating statistical and rule-based models. *Natural Language Engineering*, 23 :733–762.
[33] Lee, K., He, L., Lewis, M., and Zettlemoyer, L. S. (2017b). End-to-end neural coreference resolution. In *EMNLP*.
[34] Lu, J. and Ng, V. (2016). Event coreference resolution with multi-pass sieves. In *LREC*.
[35] Luo, X., Ittycheriah, A., Jing, H., Kambhatla, N., and Roukos, S. (2004). A mention-synchronous coreference resolution algorithm based on the bell tree. In *ACL*.
[36] Marasovic, A., Born, L., Opitz, J., and Frank, A. (2017). A mention-ranking model for abstract anaphora resolution. In *EMNLP*.
[37] McCarthy, J. F. and Lehnert, W. G. (1995). Using decision trees for coreference resolution. In *IJCAI*.
[38] Mitkov, R. (1998). Robust pronoun resolution with limited knowledge. In *COLING-ACL*.
[39] Mitkov, R. (2001). Towards a more consistent and comprehensive evaluation of anaphora resolution algorithms and systems. *Applied Artificial Intelligence*, 15 :253–276.
[40] Mitkov, R. (2007). Anaphora resolution : The state of the art. In *ACL 2007*.
[41] Mitkov, R., Evans, R., and Orasan, C. (2002). A new, fully automatic version of mitkov's knowledge-poor pronoun resolution method. In *CICLing*.
[42] Müller, M.-C. (2008). Fully automatic resolution of 'it', 'this', and 'that' in unrestricted multi-party dialog.
[43] Muzerelle, J., Lefeuvre, A., Schang, E., Antoine, J.-Y., Pelletier, A., Maurel, D., Eshkol-Taravella, I., and Villaneau, J. (2014). Ancor_centre, a large free spoken french coreference corpus : description of the resource and reliability measures. In *LREC*.
[44] Ng, V. (2003). Machine learning for coreference resolution : Recent successes and future challenges.

[45] Ng, V. (2010). Supervised noun phrase coreference research : The first fifteen years. In *ACL*.

[46] Ng, V. and Gardent, C. (2002). Improving machine learning approaches to coreference resolution. In *ACL*.

[47] Poesio, M. (2004). The mate/gnome proposals for anaphoric annotation, revisited. In *SIGDIAL Workshop*.

[48] Poesio, M. and Artstein, R. (2008). Anaphoric annotation in the arrau corpus. In *LREC*.

[49] Poesio, M., Ponzetto, S., and Versley, Y. (2011). Computational models of anaphora resolution : A survey.

[50] Ponzetto, S. P. and Strube, M. (2006). Exploiting semantic role labeling, wordnet and wikipedia for coreference resolution. In *HLT-NAACL*.

[51] Pradhan, S., Ramshaw, L. A., Weischedel, R. M., MacBride, J., and Micciulla, L. (2007). Unrestricted coreference : Identifying entities and events in ontonotes. *International Conference on Semantic Computing (ICSC 2007)*, pages 446–453.

[52] Raghunathan, K., Lee, H., Rangarajan, S., Chambers, N., Surdeanu, M., Jurafsky, D., and Manning, C. D. (2010). A multi-pass sieve for coreference resolution. In *EMNLP*.

[53] Rajagopal, D., Hovy, E. H., and Mitamura, T. (2016). Unsupervised event coreference for abstract words.

[54] Reichler-Béguelin, M.-J. (1988). Anaphore, cataphore et mémoire discursive. *Pratiques*, 57(1) :15–43.

[55] Soon, W. M., Ng, H. T., and Lim, C. Y. (2001). A machine learning approach to coreference resolution of noun phrases. *Computational Linguistics*, 27 :521–544.

[56] Steinberger, J., Poesio, M., Kabadjov, M. A., and Jezek, K. (2007). Two uses of anaphora resolution in summarization. *Inf. Process. Manage.*, 43 :1663–1680.

[57] Strube, M. and Müller, C. (2003). A machine learning approach to pronoun resolution in spoken dialogue. In *ACL*.

[58] Strube, M., Rapp, S., and Müller, C. (2002). The influence of minimum edit distance on reference resolution. In *EMNLP*.

[59] Sukthanker, R., Poria, S., Cambria, E., and Thirunavukarasu, R. (2018). Anaphora and coreference resolution : A review. *ArXiv*, abs/1805.11824.

[60] Tarrade, L., Lopez, C., Panckhurst, R., and Antoniadis, G. (2017). Typologies pour l'annotation de textes non standard en français. In *TALN*, pages 118–125.

[61] Tetreault, J. R. (2001). A corpus-based evaluation of centering and pronoun resolution. *Computational Linguistics*, 27 :507–520.

[62] Uryupina, O., Kabadjov, M. A., and Poesio, M. (2016). Detecting non-reference and non-anaphoricity. In *Anaphora Resolution - Algorithms, Resources, and Applications*.

[63] Webber, B. L. (1991). Structure and ostension in the interpretation of discourse deixis. *ArXiv*, cmp-lg/9708003.

[64] Werlen, L. M. and Popescu-Belis, A. (2017). Using coreference links to improve spanish-to-english machine translation. pages 30–40.
[65] Wiseman, S., Rush, A. M., and Shieber, S. M. (2016). Learning global features for coreference resolution. In *HLT-NAACL*.
[66] Wiseman, S., Rush, A. M., Shieber, S. M., and Weston, J. (2015). Learning anaphoricity and antecedent ranking features for coreference resolution. In *ACL*.
[67] Yang, X. and Su, J. (2007). Coreference resolution using semantic relatedness information from automatically discovered patterns. In *ACL*.
[68] Yang, X., Zhou, G., Su, J., and Tan, C. L. (2003). Coreference resolution using competition learning approach. In *ACL*.
[69] Zeldes, A. and Zhang, S. (2016). When annotation schemes change rules help : A configurable approach to coreference resolution beyond ontonotes. In *CORBON@HLT-NAACL*.

Aspects acoustiques et articulatoires des disfluences bègues dans un contexte de bilinguisme

Le cas du mina et du français

D. GBEDAHOU[1],

F. HIRSCH[1]

[1]PRAXILING, UNIVERSITÉ PAUL-VALÉRY MONTPELLIER 3,

PRAXILING

Plan

I. Généralités sur le bégaiement
II. Objectifs de l'étude
III. Méthodologie
IV. Résultats préliminaires
V. Discussion

1. Généralités sur le bégaiement

Définition

Trouble de l'élocution se traduisant par la survenue de disfluences altérant le rythme de la parole (Monfrais-Pfauwadel, 2014)

Trois principaux types de disfluences, pouvant être combinées :

- Répétitions
- Prolongations
- Blocages

PRAXILING

Types et étiologie du bégaiement

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

cnrs

Plusieurs types de bégaiement : développemental transitoire, développemental persistant, neurogène

Le *bégaiement développemental* :

- Démarre entre 2 et 7 ans
- Prévalence : entre 0,72% et 1% de la population adulte (Piérart, 2011, Yairi & Ambrose, 2013)
- Taux de rémission : 80%

Etiologie

- Origines multi-factorielles (Monfrais-Pfauwadel, 2014)
- Facteurs neurologiques (De Nil *et al.*, 2008 ; Connally *et al.*, 2014)
- Facteurs génétiques (Drayna & Kang, 2011)

PRAXILING

Phonétique du bégaiement

Sur l'anglais

Syllabes accentuées (Wingate, 1979)

Consonnes non-voisées (Jayaram, 1983, Blomgren, 2012)

Consonnes labio-dentales (Blomgren, 2012)

Consonnes constrictives (Jayaram, 1983)

Clusters consonantiques en début de mot (Wolk, 2000)

Sur le français et le slovaque

Consonnes non-voisées (Didirková, 2016)

Consonnes occlusives (Didirková, 2016)

PRAXILING

Bilinguisme et bégaiement

Fréquence du bilinguisme : environ 4000 langues parlées pour 193 pays

Bilinguisme = facteur favorisant le bégaiement ?

Manifestation du bégaiement : dans une langue ou les deux ?

Si les deux, bégaiement identique, différent ?

Le type de bilinguisme a-t-il une influence sur le bégaiement ?

PRAXILING

Bilinguisme et bégaiement (2)

Plusieurs observations contradictoires:

Bégaiement uniquement dans la langue première (Dale, 1977) ;

Présence de disfluences pathologiques dans les deux langues, avec une sévérité différente dans ces deux langues -> *difference-hypothesis* (Shenker *et al.*, 1998; Jankelowitz & Bortz, 1996; Maruthy *et al.*, 2015);

Sévérité comparable dans les deux langues pratiquées par le sujet-> *same-hypothesis* (Van Riper, 1982; Mohammadi *et al.*, 2012)

PRAXILING

Gbedahou *et al.* (2019 A et B)

Les locuteurs bégaient moins en mina (L1) qu'en français (L2)

Manifestations différentes du bégaiement en français et en mina

- En français:
  - Voyelles antérieures > voyelles postérieures
  - Voyelles nasales > voyelles orales
  - Consonnes non-voisées > consonnes voisées
- En mina:
  - Voyelles ouvertes > voyelles fermées

II. Objectif et Hypothèse

Objectif et Hypothèse

Objectif

Vérifier si les disfluences sont plus sévères d'un point de vue articulatoire et acoustique dans une langue par rapport à l'autre

Hypothèse

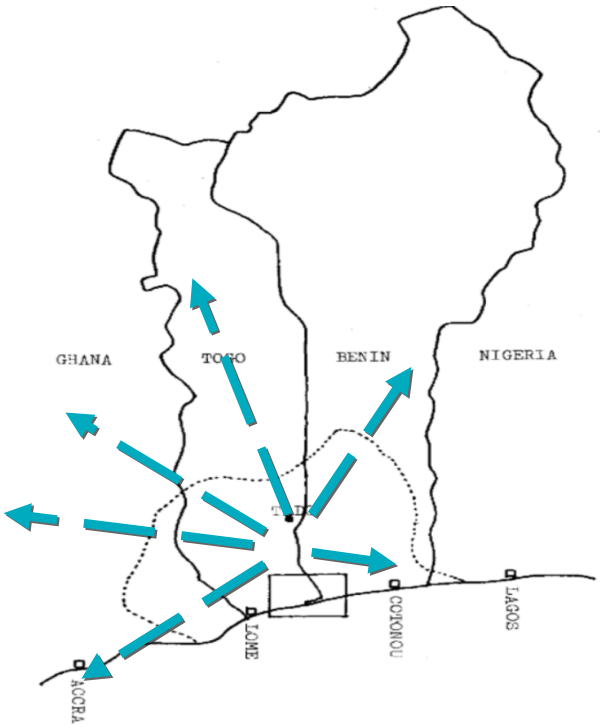
Les mouvements articulatoires seraient davantage désordonnés et anarchiques dans une langue par rapport à l'autre et leur durée ne serait pas la même en fonction de la langue.

PRAXILING

Le Mina

Expansion du Mina

Dynamisme commercial des locuteurs en particulier les « Nana-Benz »



PRAXILING

Le système phonétique du mina

	Bilabiales	Labio-dentales	Alvéolaires	Palato-alvéolaires	Rétroflexes	Palatales	Vélaires	Glottales	Labio-vélaires	
Plosives	P b		t d		ɖ	c ɟ	k g		k͡p	g͡b
Nasales	m		n		ɳ	ɲ	ŋ			
Fricatives		f v	s z				x	h		
Affriquées				d͡ʒ						
Spirantes			l						w	

Système consonantique du Mina / Adapté de Bole-Richard, 1983

Système vocalique

12 voyelles (Gblem-Poidi et Kantchoa 2012)

✓Sept voyelles orales : [a], [e], [ɛ], [i], [o], [ɔ], [u]

✓Cinq voyelles nasales : [ã], [ɔ̃], [ɛ̃], [ĩ], [ũ]

Syllabes en mina : Syllabes V, V, VC, mais également des séquences tri-consonantiques CCCV, [klw] dans [àklwí] (bouillie locale).

PRAXILING

Système tonal du mina

Deux tonèmes Haut et Bas

Deux réalisations modulées (Bas-Haut et Haut-Bas)

Sur le plan phonotactique, les combinaisons [V_1CV_2] comportent principalement un [a] ou un [e] comme voyelle initiale et que, dans pareille configuration, V1 est produite avec un ton majoritairement bas.

III. Méthodologie

Population et conditions d'acquisition des données

Population

Une locutrice

- Âgée de 27 ans
- L1 français, le mina à 4 ans
- Bégaiement développemental persistant

Lieu

- Laboratoire lorrain de recherche en informatique et ses applications (LORIA) Nancy

PRAXILING

Population et conditions d'acquisition des données

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

cnrs

Technique

Articulographe électromagnétique; (visualisation sur trois axes (X, Y et Z)

Source: Thèse Ivana Didirkova, 2016

PRAXILING

Placement des capteurs

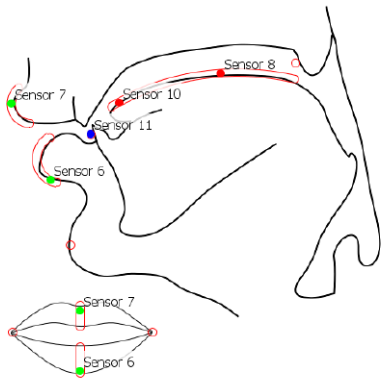
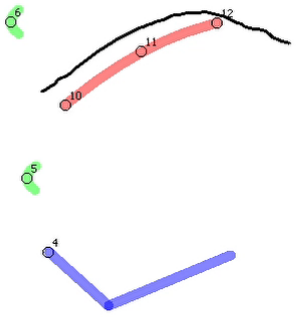
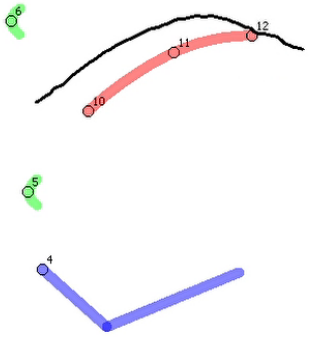


Figure 1: Placement of sensors for the mandible, the lips and the tongue.

PRAXILING

Exemples





PRAXILING

Tâches

Parole spontanée

- Questions relatives aux activités effectuées la veille + un évènement marquant dans l'actualité

Langage narratif

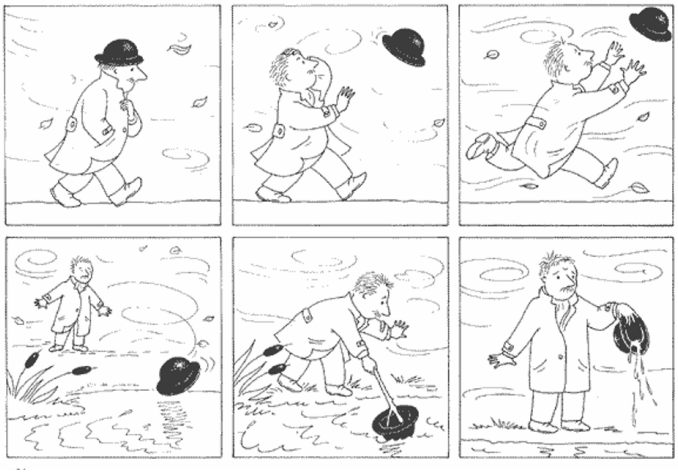
- Frog where are you? Histoire de 30 planches (Mercer Mayer)

Description d'images séquentielles

- Le sujet observe une série d'images séquentielles qu'il restitue ensuite de mémoire

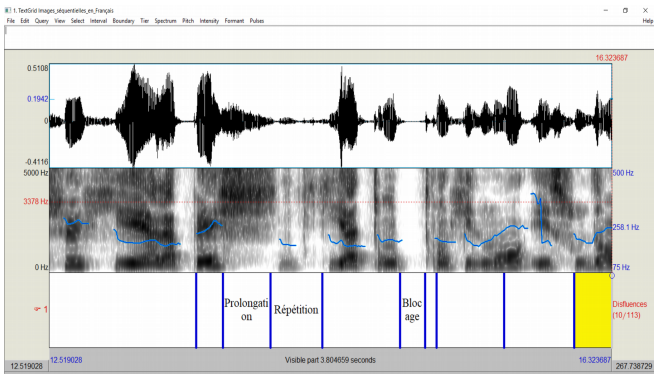
Epreuve de description d'images séquentielles

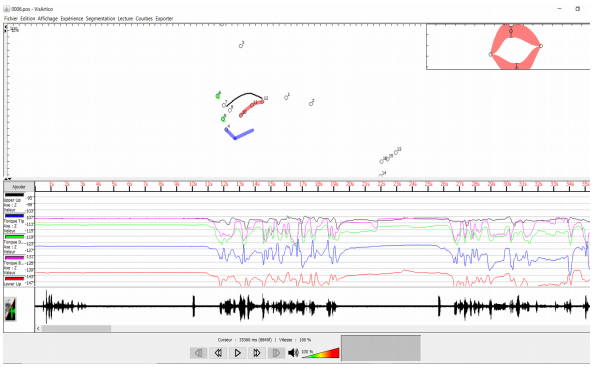
Epreuve de description d'images séquentielles



PRAXLING

Analyses





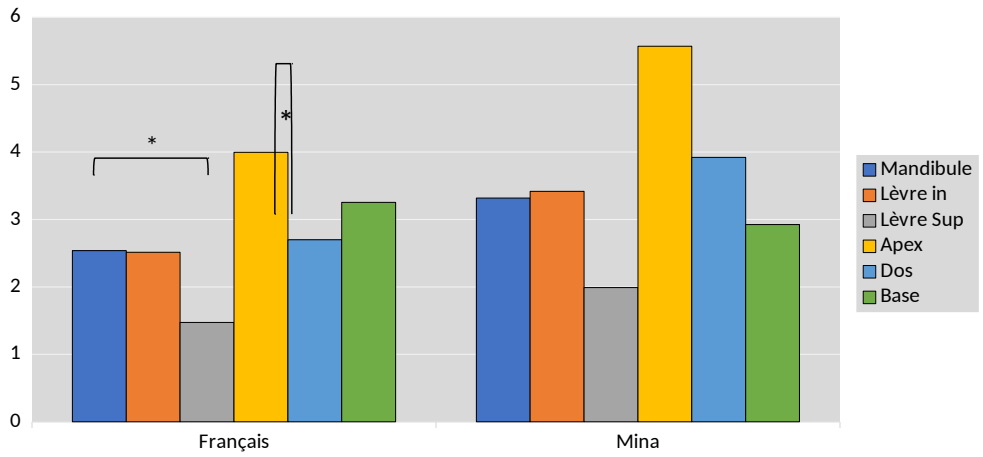
Observations

La durée des disfluences en fonction du type de disfluences et de la langue

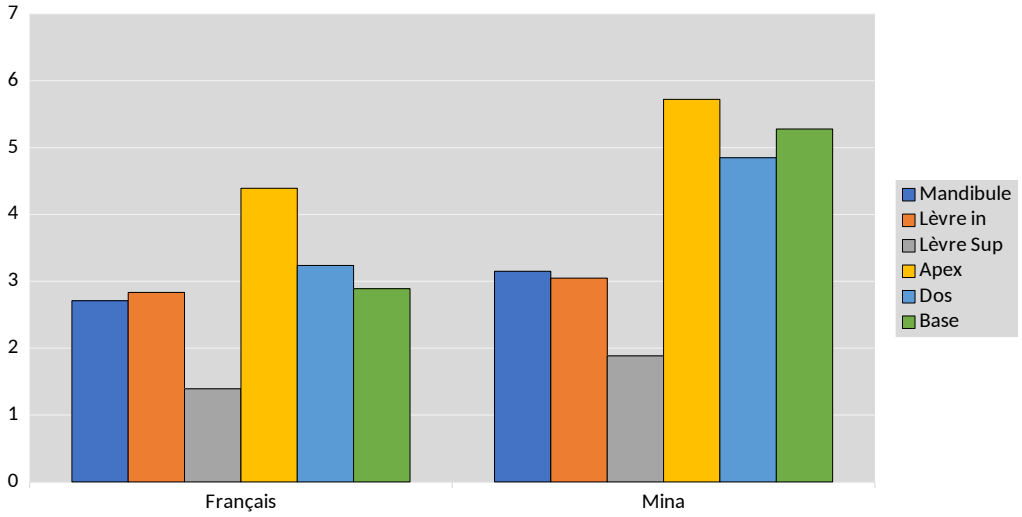
L'amplitude du mouvement des articulateurs

IV. Résultats

Mobilité des articulateurs pendant les prolongations

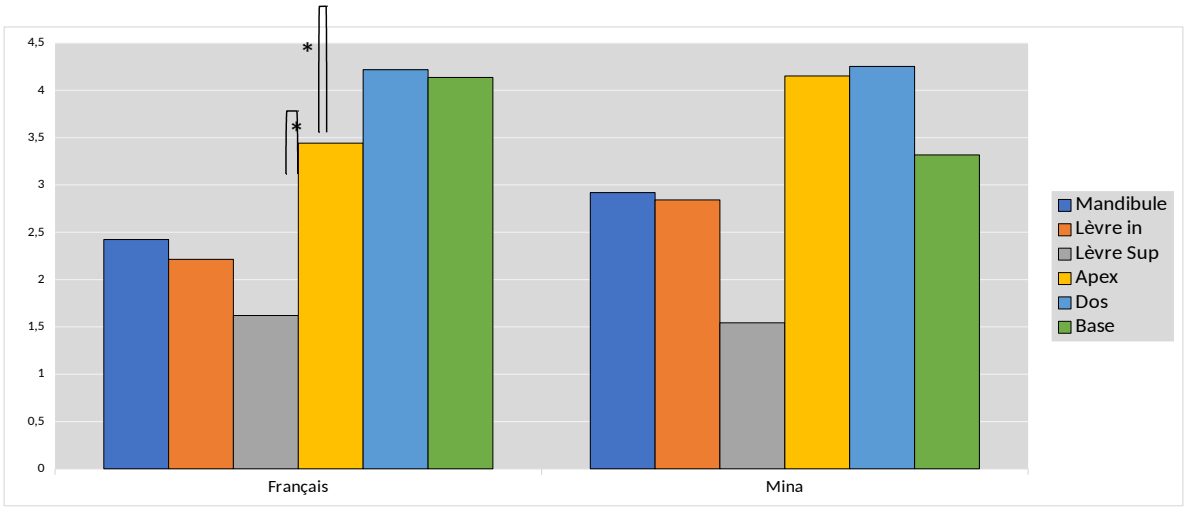


Mouvement des articulateurs pendant les répétitions



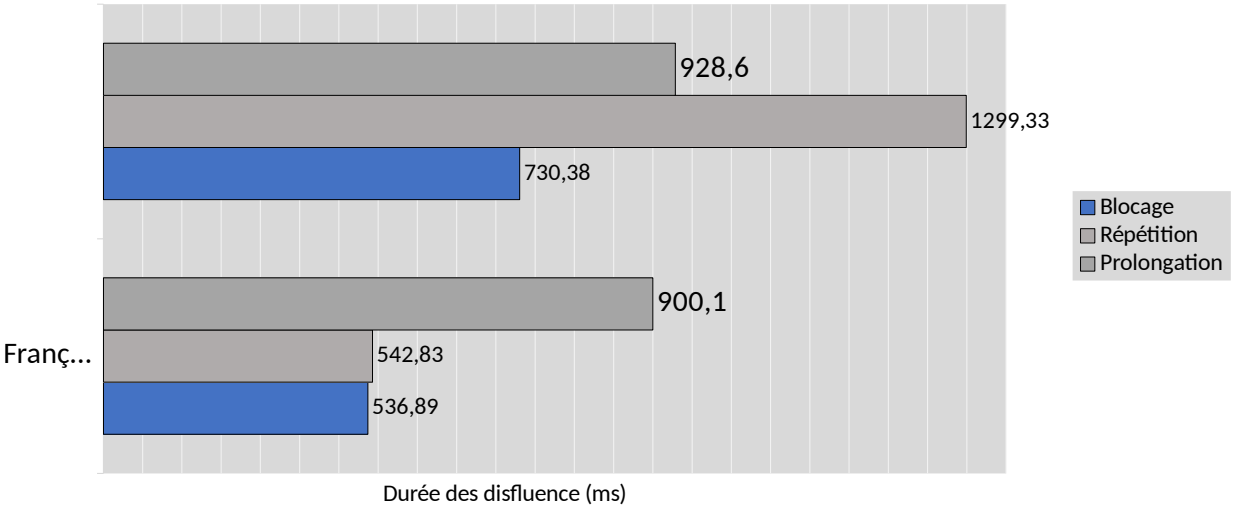
PRAXILING

Mobilité des articulateurs pendant les blocages



PRAXILING

Durée des disfluences en français etpen mina



V. Discussion

PRAXLING

Pour résumer

✓Manifestations différentes du bégaiement en français et en mina

Sur le plan articulatoire

✓Mouvements plus amples des articulateurs pendant les disfluences en langue mina (tendance?)

➢Particularités prosodiques du mina?

✓Blocage : présence de mouvements inaudibles (Didirkova, 2016)

Sur le plan acoustique: durée plus importantes en mina

Dans les deux langues : Apex plus mobile et lèvre supérieure moins

PRAX LING

Limites et perspectives

- Nombre de locuteurs à augmenter
- **Étudier les contextes phonétiques des disfluences**
- **Étudier les structures syllabiques disfluentes**
- Étudier le lien entre les traits tonals et l'apparition des disfluences
- Étudier d'autres niveaux linguistiques (prosodie par exemple)

Merci pour votre attention

REMERCIEMENTS : ECOLE D'ORTHOPHONIE DE LOMÉ, PROJET BENEPHIDIRE

1.7 La constitution d'un corpus d'ellipses erronées pour la représentation et la transmission des connaissances

Laura Noreskal-(présentation)

La constitution d'un corpus d'ellipses erronées pour la représentation et la transmission des connaissances

Laura Noreskal[1]

[1]Université Paris Nanterre, MoDyCo UMR 7114, 200 Avenue de la République, 92001 Nanterre, France
laura.noreskal@parisnanterre.fr

Résumé. Cette recherche fait partie d'un projet national, écri+[1], visant la mise en place d'un dispositif d'évaluation, de formation et de certification de la maîtrise du français. Dans cet article, nous proposons de présenter nos choix méthodologiques pour la constitution d'un grand corpus de constructions elliptiques erronées dans le but de leur détection automatique. Le corpus constitué permettra non seulement de tester différentes méthodes de traitement automatique du langage mais également d'avoir une meilleure compréhension des facteurs générateurs d'ellipses erronées chez les étudiants.

Mots-clés. Ellipse, corpus, français, traitement automatique du langage, rédaction d'étudiants

Abstract. This research is part of a national project aimed at setting up a system for evaluating, training and certifying French language proficiency, ecri+. In this article, we propose to present our methodological choices for the constitution of a large corpus of faulty elliptic constructions in order to detect them automatically. The corpus will allow not only to test different methods of automatic language processing but also to have a better understanding of the factors generating erroneous ellipsis in student writings.

Keywords. Ellipsis, corpus, French, natural language processing, student writing

[1]Numéro ANR : ANR17NCUN0015

1 Introduction

À l'ère du numérique, les pratiques pédagogiques sont modifiées par l'évolution des technologies et des usages. Face aux difficultés que rencontrent de nombreux francophones dans la maîtrise du français écrit à l'université, il semble nécessaire de tirer profit de l'innovation numérique pour déployer de nouvelles méthodes d'apprentissage et de représentations des connaissances. Dans ce but, un collectif d'universités développe actuellement un dispositif national d'évaluation, de formation et de certification des compétences d'expression et de compréhension écrites en français : écri+ (ANR17NCUN0015). Il se compose de nombreux modules médiatisés, parmi lesquels un module de rédaction de textes brefs avec détection et correction automatique d'erreurs.

Dans ce cadre, notre contribution porte plus spécifiquement sur la constitution d'un large corpus représentatif des ellipses sujettes aux erreurs dans les rédactions des étudiants. L'ellipse mobilise beaucoup de contraintes linguistiques, comme par exemple des contraintes de parallélisme entre la séquence source et la séquence ellipsée (Abeillé 2008, Desmets 2008, Bîlbîie 2011). Leur non-respect provoque immédiatement une instabilité de l'énoncé (ex : **Les enfants sont allés en Allemagne et Italie*). Parmi les zones de difficultés rédactionnelles identifiées, l'ellipse est donc particulièrement révélatrice des compétences des étudiants. Plusieurs modules ont été élaborés dans le cadre des ressources de l'UOH (Université Numérique des Humanités)[2] afin de dresser un premier repérage des contextes syntaxiques et rédactionnels sensibles favorisant l'apparition d'ellipses (coordination, juxtaposition, comparatives, etc.) et de proposer un premier classement d'occurrences reposant sur la formulation d'hypothèses sur la source des erreurs (défaut du lien logique entre propositions, défaut de la valence du verbe ellipsé, etc.). En complément à ces modules, nous avons décidé de mener une étude linguistique sur les productions correctes et erronées de constructions elliptiques. Cette recherche sur les ellipses sera la première expérimentation pour identifier les erreurs syntaxiques dans les textes brefs d'étudiants.

2http://www.uoh.fr/front; voir en particulier la plateforme http://uoh.univ-montp3.fr/j_ameliore_ma_maitrise_du_francais/portail/.

2 État de l'art

2.1 Les constructions elliptiques

L'ellipse est un phénomène très fréquent qui a été beaucoup étudié en français (Zribi-Hertz 1986 ; Busquets & Denis 2001 ; Mouret 2007 ; Abeillé 2008 ; Desmets 2008, Bîlbîie 2011). Dans un énoncé constitué de deux propositions A et B liées par une relation sémantique, une ellipse est l'omission dans B d'éléments récupérables grâce à A qui fournit un contexte direct. La proposition B est donc syntaxiquement incomplète mais reste compréhensible grâce aux informations données par les éléments réalisés dans A. La phrase elliptique est donc régie par des conditions sémantiques, pragmatiques et syntaxiques (Desmets 2008 ; Abeillé & Mouret 2008). De nombreuses recherches ont permis de distinguer plusieurs types d'ellipses dont le gapping (Ross 1970) *Pierre aime les fraises et Julie les abricots*, le sluicing (Ross 1969) *Il a réussi et j'ignore comment*, les ellipses "polaires" *Jean est parti et Pierre aussi*, les phrases à factorisation droite *Marie aime mais Pierre déteste le football* et les VP-Ellipsis ou ellipses modales *Pierre a mangé autant de bonbons qu'il a pu*.

Les recherches ont également montré que les ellipses mobilisaient beaucoup de contraintes linguistiques, notamment des contraintes de parallélisme syntaxique, sémantique et pragmatique entre la séquence source et la séquence ellipsée (Abeillé 2008, Desmets 2008, Bîlbîie 2011). Chaque élément ellipsé doit avoir un élément parallèle qui partage les mêmes caractéristiques linguistiques que lui.

(1) *Il grêle et danse.
(2) Il chante et danse.
(3) *Je vais en Espagne et Marie Allemagne.
(4) Je vais en Espagne et Marie en Allemagne.

Ainsi, dans l'exemple (1), l'élément ellipsé (pronom personnel) et l'élément réalisé (pronom impersonnel) ne partagent pas les mêmes caractéristiques linguistiques, ce qui provoque une instabilité de l'énoncé. Dans l'exemple (3), la contrainte non respectée est syntaxique. En effet, le verbe *aller* étant bivalent, il attend un sujet agentif (syntagme nominal) et un syntagme prépositionnel complément locatif. Or, en (3), les deux éléments réalisés dans la séquence ellipsée sont des syntagmes nominaux (*Marie* et *Allemagne*). Ainsi, les exemples (1) et (3) ne respectent pas la contrainte de parallélisme, ce qui empêche la compréhension de l'énoncé. Les exemples (2) et (4) sont des contextes corrects où les contraintes s'appliquent de façon satisfaisante.

2.2 La détection et la résolution automatique d'ellipses

Depuis quelques années, plusieurs recherches ont été menées sur l'annotation de corpus pour la détection et la résolution automatique d'ellipses (entre autres Nielsen

2005 ; Bos & Spenader 2011 ; Gandón-Chapela 2017). Elles ont principalement cherché à détecter et à résoudre des VP-Ellipsis dans des corpus anglophones tels que le British National Corpus (BNC) et le Wall Street Journal (WSJ). Mais aucune n'a porté sur la correction d'ellipses erronées ou sur les productions écrites des étudiants francophones, et c'est sur cette question spécifique que porte notre travail. Cet article expose donc l'étape primordiale de constitution d'un corpus spécifique qui permettra de tester différentes méthodes d'apprentissage automatique pour automatiser la détection et la correction des ellipses.

3 Constitution du corpus

3.1 Méthodologie de constitution du corpus

Afin de constituer notre corpus d'ellipses erronées, nous devons répondre à deux problématiques. La première concerne les types de rédaction privilégiant l'apparition d'ellipses erronées. Nous cherchons à constituer un corpus d'erreurs d'une taille considérable afin de pouvoir comprendre le phénomène et l'analyser. De ce fait, il est nécessaire de savoir dans quel type de productions écrites des étudiants les erreurs sont le plus présentes. Pour répondre à cette question, nous avons collecté plusieurs rédactions d'étudiants de différents niveaux et domaines d'études tels que Science du Langage, Histoire, LLCER Anglais, Droit et Économie. Nous avons ensuite classé les rédactions en deux catégories : les rédactions spontanées (partiels et exercices faits en classe) et les rédactions préparées (devoirs maison, rapports de stage et mémoires). En catégorisant les rédactions, nous pensons pouvoir déterminer la probabilité d'apparition d'ellipses erronées selon la nature de la production écrite. Les rédactions spontanées n'étant pas soumises aux relectures, nous pensons qu'elles contiennent plus d'erreurs que les rédactions préparées. Afin de valider notre hypothèse, nous avons recueilli les deux catégories de rédaction.

Néanmoins, déterminer quel type de rédaction est le plus susceptible de contenir des erreurs ne suffit pas. Il est nécessaire de savoir également quel type d'ellipse est le plus sujet aux erreurs. Notre travail implique de comprendre au mieux le phénomène afin de produire, au terme de notre étude, un outil capable de détecter les ellipses erronées. Il nous faut donc pouvoir établir une typologie des erreurs d'ellipses dans les rédactions. Nous avons alors décidé d'extraire les ellipses erronées avec un contexte conséquent : un contexte gauche de 2 à 3 phrases et un contexte droit d'une phrase. L'ellipse est un phénomène qui ne se produit généralement pas au-delà de deux phrases mais nous avons considéré que le contexte ne devait pas être négligé car il pourrait nous donner des informations sur les causes de l'erreur. Par ailleurs, chaque rédaction contenant une ellipse erronée a été gardée dans son intégralité dans le but de garder des informations comme le placement de l'ellipse dans le texte de rédaction.

Ainsi, depuis le début de notre recherche, nous avons collecté 250 rédactions dont 82 devoirs maison, 1 mémoire, 7 rapports de stage, 119 partiels et 41 exercices faits en classe.

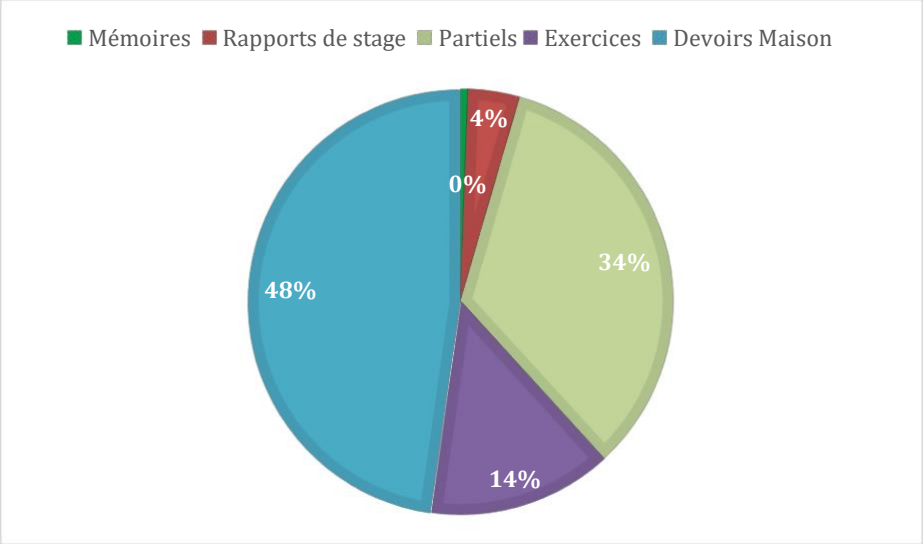
Table 1. Exemples d'ellipses erronées du corpus

Rédaction	Type de rédaction	Niveau	Domaine
Il s'agit soit d'un couple hétérosexuel stérile ou ayant des problèmes de fertilité, soit **<err>** un couple homosexuel ne pouvant, par nature, pas avoir d'enfant.	Exercice	Licence 1	LLCER Anglais
[...] La référence est générique car on parle de toutes les salades et pas **<err>** une en particulier […]	Partiel	Licence 1	Sciences du langage
[…] Toutes ces dispositions permettent à l'élève de réussir et donc **<err>** essayer d'augmenter l'égalité des chances […]	Devoir maison	Master 1	MEEF

3.2 Premières observations

Après avoir récupéré ces 250 rédactions d'étudiants, nous avons remarqué que les ellipses de type gapping posent le plus de problèmes. Nous avons pu extraire 178 constructions elliptiques erronées dont 25 des exercices faits en classe (14%), 1 du mémoire (0%), 7 des rapports de stage (4%), 60 des partiels (34%) et 85 des devoirs maison (48%) (voir Table 2).

Table 2. Répartition en pourcentage des erreurs d'ellipses



Au début de notre recherche, nous avions émis l'hypothèse que les rédactions spontanées (partiels et exercices) contiendraient plus d'erreurs que les rédactions préparées. Or, nous avons pu remarquer que les devoirs maison, que nous avons considérés comme des rédactions préparées, sont les productions qui contiennent paradoxalement le plus d'erreurs dans le corpus constitué à ce jour. Il s'agit d'un point qui nécessite de plus amples réflexions. Les devoirs maison, bien que préparés, ne sont peut-être pas sujets à la même considération de la part des étudiants que les rapports de stage ou les mémoires. Est-ce parce que l'enjeu du travail n'est pas le même pour les devoirs maison que la qualité de l'écrit est moins surveillée ? De fait, les erreurs sont donc beaucoup plus présentes dans ces devoirs que dans les autres rédactions préparées.

Toutefois, comme attendu, les résultats obtenus sur cet échantillon de montrent que le nombre d'erreurs dans les rédactions spontanées est également important (48%). Les résultats confirment donc que les rédactions les plus sujettes aux erreurs sont celles qui ne sont pas obligatoirement soumises à l'étape de relecture. Ainsi, les mémoires (0,6%) et les rapports de stage (4%) sont les rédactions qui contiennent le moins d'erreurs malgré leurs tailles, ce qui rend le relevé d'erreurs peu rentable sur ce type de données.

4 Conclusion et perspectives

Dans cet article, nous avons exposé les différentes études menées sur les constructions elliptiques puis nous avons détaillé les choix méthodologiques pour la constitution de notre corpus dans le but de créer un outil de détection automatique des ellipses dans les copies des étudiants. Nous avons donc choisi de collecter tous les types de rédactions d'étudiants afin d'observer quelles étaient les rédactions les plus sujettes aux erreurs. De plus, nous avons également cherché à savoir quel était le type d'ellipse

le plus susceptible d'être erroné. Les premières observations montrent que les constructions elliptiques erronées sont majoritairement présentes dans les rédactions sans relecture et que le gapping est le principal phénomène touché par une mauvaise construction.

En perspective, nous comptons continuer à collecter davantage de données afin d'atteindre un volume plus important (plus de 200 erreurs) dans le but de tester plusieurs méthodes de traitement automatique du langage pour la détection automatique. Un large corpus permettra également d'analyser plus en détails les caractéristiques propres aux rédactions des étudiants et ainsi permettre une meilleure compréhension des problèmes d'ellipses et de constructions qu'ils rencontrent lors des productions écrites. Notre étude donnera lieu alors à de nouvelles méthodes de travail et de formation pour aider les étudiants à maîtriser leur langue maternelle lors de leurs études universitaires.

Références

1. Abeillé A., Bilbiie G., & Mouret F. : Gapping in Romance : a fragment analysis, *International Conference on Elliptical constructions*, LLF & Université Paris Diderot, Chicago Center, Paris (2008)
2. Abeillé A., Mouret F. : Quelques contraintes sur les coordinations elliptiques en français. *Revue de sémantique et de pragmatique* (2008)
3. Bédard M., Tissot I. : L'Ellipse du complément du nom dans une coordination à deux ellipses. Actes du *Xe Colloque des étudiants en sciences du langage* (2006) 109-134
4. Bilbiie G. : Grammaire des constructions elliptiques. Une étude comparative des phrases sans verbe en roumain et en français, *Thèse de Doctorat*, Université Paris 7 (2011)
5. Bos J., Spenader J.: An annotated corpus for the analysis of VP Ellipsis. *Language Resources and Evaluation* 45 (2011) 463-494
6. Bourreau P. : Traitements d'ellipses : deux approches par les grammaires catégorielles abstraites, Actes de la *20e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2013*) pp.215-228
7. Busquets J., Denis P. : L'Ellipse modale en français. *Cahiers de Grammaire* 26 (2001) 55-74
8. Dagnac A. : L'Ellipse modale en français : argument pour une ellipse du TP. *Congrès Mondial de Linguistique Française* - CMLF'08 (2008)
9. Desmets M. : L'ellipse dans les constructions en comme, *LINX* (2008)
10. Gandón-Chapela1 E.: Hunting for Post-Auxiliary Ellipsis in a parsed corpus of English. *Research in Corpus Linguistics 4* (2016) 33-38
11. Hardt D.: An empirical approach to VP Ellipsis. *Computational Linguistics* 23(4) (1997) 525–541
12. Ingham R., Roger G. : L'Ellipse verbale en français : point de vue diachronique. Actes du *XXVIIe Congrès international de linguistique et de philologie romanes*. (2013)

13. Kertz L., Kehler A., Elman J. L.: Evaluating a Coherence-Based Model of Pronoun Interpretation. *Ambiguity in Anaphora Workshop Proceedings* (2006) 49-56
14. Kiefer M., Schuler S., Mayer C., Trumpp N. M., Hille K., & Sachse S.: Handwriting or Typewriting? The Influence of Pen- or Keyboard-Based Writing Training on Reading and Writing Performance in Preschool Children. *Advances in cognitive psychology, 11(4)* (2015) 136-46. doi:10.5709/acp-0178-7
15. Mouret F. : Grammaire des constructions coordonnées en français, *Thèse de Doctorat*, Université Paris 7 (2007)
16. Nielsen L.: A corpus-based study of verb phrase Ellipsis identification and resolution. *Thèse de Doctorat*. King's College London (2005)
17. Troia G., Harbaugh, A., Shankland R., Wolbers K., & Lawrence, A.: Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing.* 26(1) (2013) 17-44
18. Ross J. R.: Gapping and the order of constituents. In Manfred Bierwisch & Karl Erich Heidolph (eds.), *Progress in linguistics* (1970) 249-259
19. Ross, J. R.: Guess who? In Robert Binnick, Alice Davison, Georgia Green, and Jerry Morgan (eds.), Papers from the *5th regional meeting of the Chicago Linguistic Society*, 252-286. Chicago Linguistic Society : Chicago, Ill (1969)
20. Zribi-Hertz A. : Relations anaphoriques en français. Thèse de Doctorat d'État, Université Paris 8 (1986)

La constitution d'un corpus d'ellipses erronées pour la représentation et la transmission des connaissances

Laura NORESKAL
Doctorante en linguistique – traitement automatique du langage
Laboratoire MODYCO – UMR 7114 & Université Paris Nanterre
laura.noreskal@parisnanterre.Fr

- Projet national impliquant 13 universités françaises
- Utiliser le numérique pour développer de nouvelles formes d'évaluation et de formation
- Développement d'un dispositif national d'évaluation, de formation et de certification des compétences d'expression et de compréhension écrites en français
- Plusieurs modules médiatisées déjà déployés notamment à Paris Nanterre

2

Les constructions elliptiques

- Premier repérage des zones de difficultés rédactionnelles : Les constructions elliptiques sont souvent sujettes aux erreurs dans les rédactions des étudiants
- Plusieurs modules pour déterminer les contextes syntaxiques favorisant l'ellipse et plus exactement les erreurs
- Étude linguistique pour la détection automatique

3

La détection automatique des ellipses

Objectifs :

- Détecter automatiquement les ellipses dans des textes brefs d'étudiants
- Reconnaître les erreurs de constructions
- Proposer un feed-back

4

Les ellipses

- Dans un énoncé constitué de deux propositions A et B liées par une relation sémantique, une ellipse est l'omission dans B d'éléments récupérables grâce à A qui fournit un contexte direct. La proposition B est compréhensible grâce aux informations données par les éléments réalisés dans A.

- Selon Abeillé & Mouret (2008) : « On parle d'ellipse [dans les coordinations elliptiques] lorsque l'interprétation d'une forme syntaxique requiert plus que ce qui est fourni par les éléments qui la composent et que le matériel nécessaire pour obtenir cette interprétation est récupérable dans le contexte immédiat. »

- Plusieurs types d'ellipses ont été étudiés tels que :
    - Le gapping (Ross 1970) : Jean boit du thé et Marie du café
    - Les ellipses polaires : Marie viendra et Pierre aussi

5

Les ellipses (2)

- Les recherches ont montré que les ellipses répondent à des contraintes linguistiques qui assurent la stabilité d'un énoncé. Le non-respect de ces contraintes entraîne une instabilité et donc une incompréhension de l'énoncé. (Abeillé 2008, Desmets 2008, Bîlbîe 2011)

- « L'ellipse est un phénomène d'interface qui implique des contraintes de parallélisme syntaxique, sémantique et pragmatique » (Desmets, 2008)

(1) *Ces références dépendent du contexte de la phrase mais aussi la sémantique.
(2) Ces références dépendent du contexte de la phrase mais aussi de la sémantique.

6

La constitution d'un corpus d'erreurs d'ellipses

Objectifs :

- Repérer le type d'erreurs d'ellipses le plus fréquent dans les rédactions d'étudiants
- Déterminer le type de rédactions le plus sujet aux erreurs
- Constituer un corpus d'entrainement pour notre outil de détection automatique

7

Méthodologie

Collecte de plusieurs rédactions d'étudiants

Catégorisation des rédactions

Extraction des erreurs d'ellipse

Choix des métadonnées à garder

8

Collecte des rédactions d'étudiants

- Examens
- Exercices
- Devoirs maison
- Rapports de stage
- Mémoires

9

Catégorisation des rédactions

Rédactions spontanées :

- Examens
- Exercices

→ Temps limité, peu de relecture

Rédactions préparées :

- Devoirs maison
- Rapports de stage
- Mémoires

→ Temps de préparation plus long et possibilité de relecture par l'étudiant ou par un tiers (proche, professeur, directeur de recherches…)

10

Extraction des erreurs d'ellipse

- Lecture des rédactions dans leur totalité
- Extraction manuelle des erreurs avec un contexte gauche de 2 à 3 phrases et un contexte droit d'une phrase

11

Les métadonnées

Pour l'étude du phénomène, nous avons gardé :

- La nature de la rédaction
- Le niveau d'étude
- Le domaine d'étude
- La rédaction dans sa totalité

12

Corpus

Extraits	Type de rédaction	Niveau	Domaine
[…] Il s'agit soit d'un couple hétérosexuel stérile ou ayant des problèmes de fertilité, soit <err> un couple homosexuel ne pouvant, par nature, pas avoir d'enfant.	Exercice	L1	LLCER Anglais
[…] Le client a été blessé et n'a pas eu de soins. Il a reçu des coups de matraques, l'interdiction de faire ces besoins et <err> privé de nourriture.	Rapport de stage	M2	Droit
[…] La référence est générique car on parle de toutes les salades et pas <err> une en particulier.	Examen	L1	Sciences du langage
[…] Toutes ces dispositions permettent à l'élève de réussir et donc <err> essayer d'augmenter l'égalité des chances.	Devoir maison	M1	MEEF

13

Premières observations

Nous avons collecté 250 productions dont 178 erreurs d'ellipse :

Type de rédactions	Nombre de redactions	Taux d'erreurs
Rédactions spontanées	160	85 (47,8%)
Rédactions préparées	90	93 (52,2 %)
Total	250	178

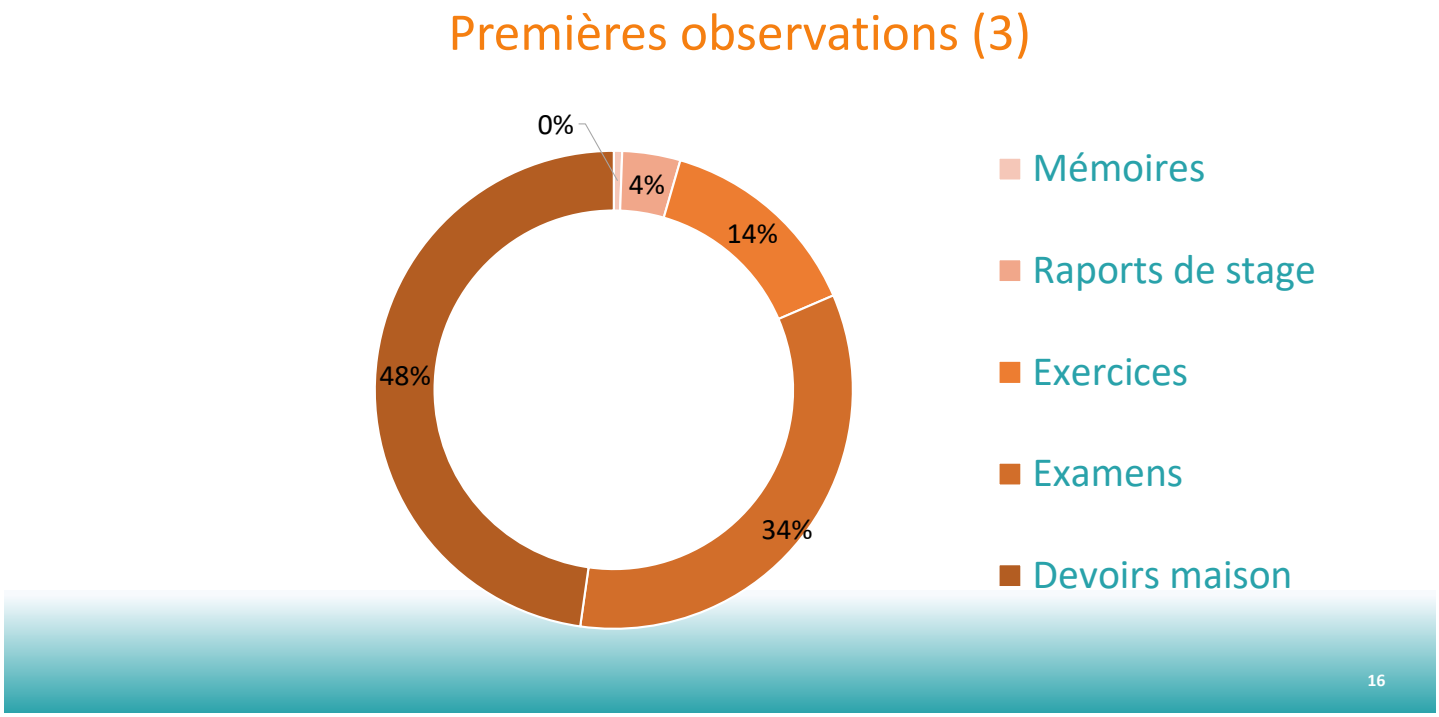
14

Premières observations (2)

- **Rédactions spontanées**: 160 productions dont 85 erreurs (**47,8 %)**
  - Examens : 119 productions dont 60 erreurs
  - Exercices : 41 productions dont 25 erreurs

- **Rédactions préparées** : 90 productions dont 93 erreurs (**52,2 %**)
  - Devoirs maison : 82 productions dont 85 erreurs
  - Rapports de stage : 7 productions dont 7 erreurs
  - Mémoire : 1 production dont 1 erreur

15


Premières observations (3)
0%
4%
14%
48%
34%
Mémoires
Raports de stage
Exercices
Examens
Devoirs maison
16

Perspectives

- ✓Constituer un corpus de 250 à 300 erreurs d'ellipses et d'ellipses bien formées
- ✓Mise en place d'un test sur 250 étudiants « provoquant » la création d'ellipses
- ✓Analyser le phénomène en prenant en compte le contexte (les verbes impliqués, la place de l'erreurs dans la rédaction…)
- ✓Tester les différentes méthodes du traitement automatique du langage

17

Merci !

References

- Abeillé A., Mouret F. : Quelques contraintes sur les coordinations elliptiques en français. Revue de sémantique et de pragmatique (2008)
- Bilbiie G. : Grammaire des constructions elliptiques. Une étude comparative des phrases sans verbe en roumain et en français, Thèse de Doctorat, Université Paris 7 (2011)
- Desmets M. : L'ellipse dans les constructions en comme, LINX (2008)
- Mouret F. : Grammaire des constructions coordonnées en français, Thèse de Doctorat, Université Paris 7 (2007)
- Ross J. R.: Gapping and the order of constituents. In Manfred Bierwisch & Karl Erich Heidolph (eds.), Progress in linguistics (1970) 249-259

1.8 Aspects acoustique et articulatoire du bégaiement dans un contexte bilingue : cas des locuteurs mina-français

Dodji Gbedahou - (présentation)

Aspects acoustiques et articulatoires des disfluences bègues auprès des locuteurs bilingues mina et français

Dodji Gbedahou[1]

[1] Université Paul-Valéry Montpellier 3, laboratoire Praxiling, CNRS UMR 5267, route de Mende, Montpellier, France
Dodji.gbedahou@univ-montp3.fr

Résumé. Selon certains travaux portant sur les causes du bégaiement, ce trouble se manifesterait beaucoup plus auprès des populations bilingues. Une première étude que nous avons menée sur des locuteurs bilingues mina/français laisse penser que le taux de bégayages diffère dans les deux langues. Cette proposition s'inscrit dans la continuité de cette étude et son but est d'analyser les caractéristiques articulatoires et acoustiques des disfluences dans chacune des deux langues citées. L'hypothèse est que les mouvements articulatoires seraient d'avantage désordonnés et anarchiques dans une langue par rapport à l'autre.

Mots-clés. Disfluences, bégaiement, bilinguisme, mina, français

1 Introduction

1.1. Le bégaiement

Le bégaiement peut être défini comme un trouble de la communication se traduisant par la présence de disfluences qui altèrent le rythme de la parole. Le bégaiement développemental persistant, l'une des formes que peut prendre le bégaiement, toucherait entre 0,7 et 1% (Yairi & Ambrose, 2013 ; Piérart, 2011). Les travaux portant sur l'étiologie du bégaiement laissent penser qu'il s'agit d'un trouble multi-factoriel (Monfrais-Pfauwadel, 2014), mêlant notamment des facteurs neurologiques (De Nil et al. 2008 ; Connally et al., 2014) et génétiques (Drayna et Kang, 2011). Par ailleurs, de nombreuses questions se posent quant à l'implication des pratiques linguistiques dans l'installation du bégaiement. Selon Van Borsel et al. (2001), le bégaiement se manifesterait plus auprès des populations bilingues que monolingues ; une affirmation différente de celle de Howell et al. (2009) qui observent une incidence identique dans les deux langues.

Une récente étude (Gbedahou, Didirková & Hirsch, 2019), se fondant sur des enregistrements de locuteurs bilingues mina – français, ont montré un effet de langue sur le bégaiement, ce trouble ne se manifestant pas de la même façon en mina et en français. En effet, davantage de disfluences étaient présentes dans la seconde langue citée. Cette proposition se situe dans la continuité de cette précédente recherche en s'intéressant plus précisément aux caractéristiques phonétiques des disfluences dans

chacune des deux langues. Notre objectif est de vérifier si les disfluences sont plus sévères d'un point de vue articulatoire et acoustique dans une langue par rapport à l'autre. Dans la suite de notre première étude, notre hypothèse est que les mouvements articulatoires seraient davantage désordonnés et anarchiques dans une langue par rapport à l'autre et que leur durée ne serait pas la même en fonction de la langue.

Pour mener à bien cette recherche, des données articulatoires ont été recueillies au LORIA (Laboratoire de Recherches en Informatique et ses applications, Nancy) auprès d'une locutrice bilingue (mina-français) à l'aide d'un articulographe électromagnétique. Les données obtenues ont été transcrites dans le logiciel Praat (Boersma & Weenink, 2017) et les mouvements articulatoires ont été mesurés avec le programme Visartico (Ouni et al 2012).

1.2. Quelques considérations sur le mina

Le mina est une langue parlée au sud du Togo et du Benin. Autrefois réservé au peuple guin, cette langue est devenue aujourd'hui la première langue véhiculaire dans la sous-région. Notons que le mina appartient à la catégorie des langue kwa, et au sous-groupe Gbe, qui renvoie à l'ensemble des langues dont le territoire s'étend du fleuve volta au fleuve Ouémé puis de l'océan atlantique jusqu'au-dessous de la latitude 9° Nord (Capo, Hounkpatin C, 1983). Signalons que la localisation telle que décrite a quelque peu évolué puisque le mina est utilisé par différents groupe comme langue seconde. Il s'agit notamment des éwé, adja, ouatchi, fon, gun, yorouba, adja, et pla.

Le système phonétique du mina est composé de 26 consonnes et de 12 voyelles. A la différence du français, cette langue présente des sons à double articulation et des consonnes rétroflexes. Du point de vue vocalique, il est noté la présence de quatre tons différents (Gblem-Poidi et Kantchoa, 2012).

Table 1 : Système consonantique du mina, adapté de Adapté de Bole-Richard, 1983

	Bilabiales	Labio-dentales	Alvéolaires	Palato-alvéolaires	Rétroflexes	Palatales	Vélaires	Glottales	Labio-vélaires
Plosives	P b		t d		ɖ	c	k g		k͡p g͡b
Nasales	m		n		ɳ	ɲ	ŋ		
Fricatives		f v	s z				x	h	
Affriquées				d͡ʒ					
Spirantes			l						w

2 Matériel et méthode

Le sujet ayant fait objet de cette étude a été soumis à une épreuve de description d'images séquentielles et à une activité de langage spontané.
Les données ont été acquises avec la collaboration Slim Oumi au Laboratoire Lorrain de Recherche en Informatique et ses Applications (Loria) de Nancy à l'aide d'un articulographe électromagnétique (Modèle AG 501 3D, marque Cartens).

L'articulographe est une technique utilisée dans les recherches en neurologie, phoniatrie, dentisterie, phonétique. Il permet d'obtenir des informations relatives aux mouvements des articulateurs (lèvres, langue, mâchoire, etc.) lors de la production de parole.
L'intérêt de cette technique est qu'elle permet de visualiser et de mesurer les mouvements articulatoires sur 3 axes : horizontal, vertical et latéral (X, Y et Z) ainsi que des angles PHI et THETA.

L'utilisation d'un articulographe demande la mise en place de capteurs sur les différents organes impliqués dans la parole. Ces capteurs fournissent donc des informations sur chacun des mouvements des articulateurs.
Dans cette étude, deux capteurs ont été placés sur la mastoïde, derrière les deux oreilles, un sur le front, un capteur sur chacune des lèvres, un capteur sur chacune des commissures labiales, trois capteurs sur la langue (apex, dos de la langue, racine de la langue) et un dernier sur le menton. Le sujet est assis sur une chaise placée sous une machine génératrice d'un champ électromagnétique qui induit un courant alternatif dans les capteurs, ce qui permet de mesurer la distance des capteurs par rapport aux transmetteurs.

Dans cette étude, uniquement les capteurs placés sur la langue, les lèvres, les commissures, la langue et le menton seront observés. La position des différents

capteurs sera calculée et enregistrée dans l'espace toutes les 5 millisecondes et avec une précision de 0,3 millimètres.

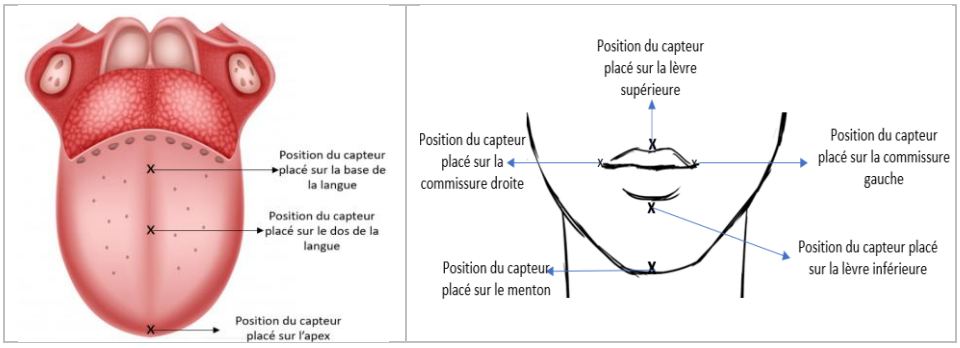


Fig 1. Mise en place des capteurs, trois capteurs placés sur la langue, quatre placés respectivement sur la lèvre supérieure, la lèvre inférieure, le côté gauche et droit des lèvres et un dernier sur le menton

Les fichiers (description d'images et langage spontané) en mina et en français ont été traités sous Praat. Les différents mots ayant fait l'objet d'une disfluence ont été segmentés. Deux lignes de segmentations ont été établies : la première est celle de la transcription phonétique des énoncés et la deuxième, celle du type de disfluence en question. Pour les transcriptions, les caractères de l'Alphabet Phonétique Internationale (API) ont été utilisés.

Pour déterminer la durée des disfluences, seule la deuxième ligne de transcription a été utilisée. En effet, l'objectif ici est d'étudier l'effet de la durée pour chacun des accidents de parole. Pour exporter la durée des disfluences, le script « Praat script » (réalisé par Cédric Gendrot) a été utilisé.

Le deuxième traitement consiste à mettre en évidence l'amplitude du mouvement des articulateurs pendant les disfluences. Pour cela les fichiers « pos », dans lesquels se situent les données sur les mouvements des articulateurs, ont été ouverts avec « Visartico ». Ensuite les segmentations préalablement effectuées sous le logiciel « Praat » ont été importées dans « Visartico ».

Une fois les données exportées dans Excel, pour chaque articulateur et pour chaque disfluence, les différences entre la valeur la plus petite et la valeur la plus grande ont été calculées. Cette valeur définie la mobilité de l'articulateur définie par Δz.

$$\Delta z = |Zn - Z1|$$ Zn étant la valeur la plus élevée et Z1 la plus petite

Dans chaque langue, l'amplitude des mouvements a été comparée pour chaque articulateur. Vu qu'il s'agit d'une analyse de plusieurs moyenne, un test de Tukey a été utilisé sous le logiciel R. Ce test permet de comparer plusieurs moyennes deux à deux à la différence des T de Student qui se limitent à deux variables. Et pour chaque comparaison, il définit une P-value.

3 Résultats

Comme mentionné, plusieurs tests ont été utilisés pour évaluer la locutrice. Néanmoins, uniquement les données de l'épreuve de description d'image seront présentées dans cet article.

3.1 Etude comparative de la mobilité des articulateurs pendant les prolongations en français et en mina

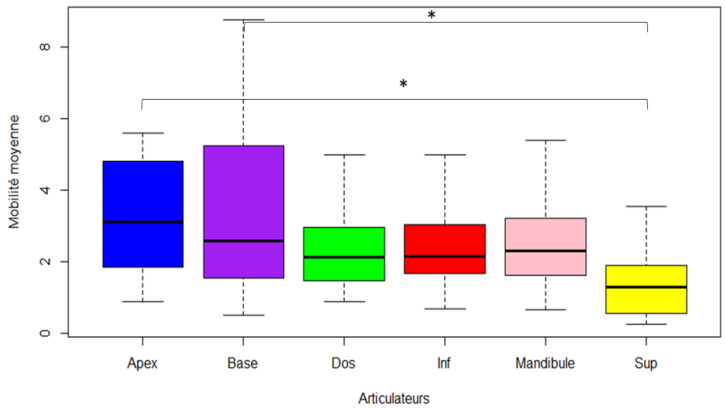


Fig.2. Mobilité des articulateurs pendant les prolongations en français. En ordonné (*Mobilité moyenne)* et en abscisse (*les articulateurs : Apex, base de la langue, dos de la langue, lèvre inférieure, mandibule, et la lèvre supérieure*). Les mouvements de l'apex et la base de la langue sont beaucoup plus exagérés que celles de la lèvre supérieure.

La figure 2 traduit le mouvement des articulateurs lors des prolongations en description d'images séquentielles. En ordonnée, nous avons la valeur moyenne des déplacements et en abscisse les différents articulateurs. On note que l'amplitude des déplacements de l'apex de la langue est la plus importante. L'écart moyen entre les valeurs les plus hautes et les valeurs les plus petites est de 3.996 mm.

Ensuite viennent la base et le dos de la langue avec un écart moyen qui est respectivement de 3.255 mm et 2.699 mm. La lèvre supérieure reste la moins mobile pendant les prolongations en français avec un écart moyen de 1.473 ms.

L'analyse statistique avec le test de Tukey permet de noter une différence significative entre les valeurs moyennes de l'apex et de la lèvre supérieure (P=0.0003). Par ailleurs on note également que les mouvements de la base de la langue sont significativement plus importants que ceux de la lèvre supérieure (P=0.0346).

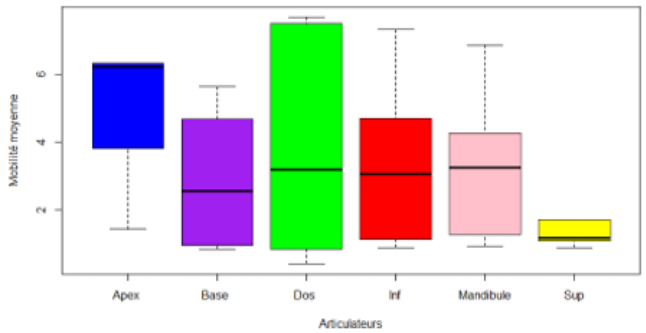


Fig.3. Mobilité des articulateurs pendant les prolongations en mina. En ordonné (*Mobilité moyenne*) et en abscisse (*les articulateurs : Apex, base de la langue, dos de la langue, lèvre inférieure, mandibule, et la lèvre supérieure*). Pas de différence significative entre les valeurs moyennes.

La figure 3 traduit le mouvement des articulateurs pendant les prolongations en mina. Les résultats permettent de noter que les mouvements de l'apex et du dos de la langue sont plus importants (5,57 mm pour l'apex et 3,92 mm pour le dos de la langue). La lèvre supérieure est l'organe la moins mobile pendant les prolongations en mina.

En résumé, pendant les prolongations, les mouvements des articulateurs sont plus importants en mina qu'en français. De plus l'apex et la base de la langue effectuent des mouvements significativement plus importants que les autres articulateurs en français ; ce qui n'est pas le cas en mina.

3.2 Mouvement des articulateurs pendant les répétitions

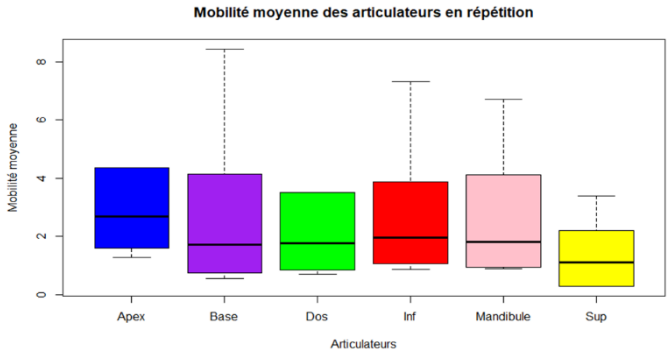


Fig.4. Mobilité des articulateurs pendant les répétitions en français. En ordonné (*Mobilité moyenne*) et en abscisse (*les articulateurs : Apex, base de la langue, dos de la langue, lèvre inférieure, mandibule, et la lèvre supérieure*). Pas de différence significative entre les valeurs moyennes.

Il est à noter dans cette étude comparative que les mouvements de l'apex bougent le plus pendant les répétitions (4.391 mm) et la lèvre supérieure est la moins mobile (1.391 mm). Néanmoins, les différences ne sont pas significatives selon les tests statistiques.

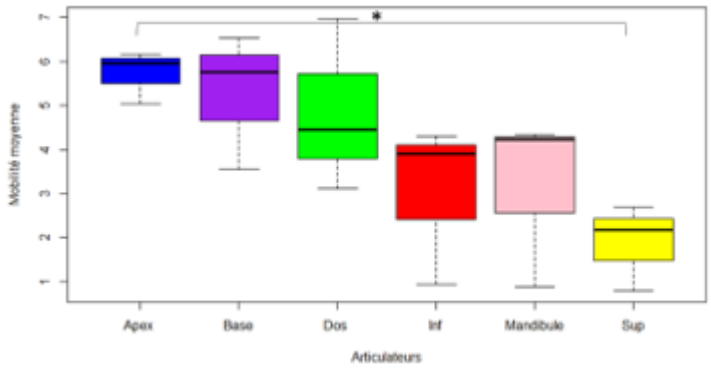


Fig.5. Mobilité des articulateurs pendant les répétitions en mina. En ordonné (*Mobilité moyenne*) et en abscisse (*les articulateurs : Apex, base de la langue, dos de la langue, lèvre inférieure, mandibule, et la lèvre supérieure*). Différence significative entre le mouvement de l'apex et celui de la lèvre supérieure

3.3 Mouvement des articulateurs pendant les blocages

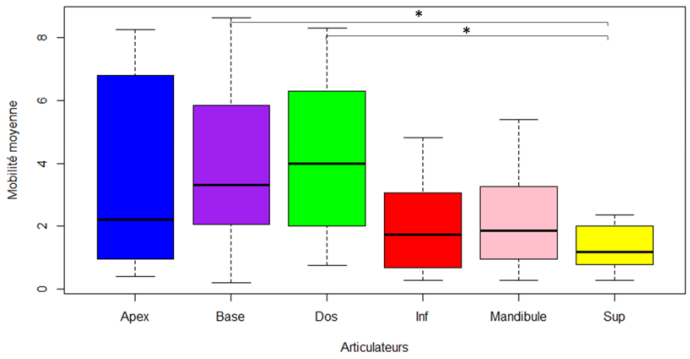


Fig.5. Mobilité des articulateurs pendant les blocages en français. En ordonné (Mobilité moyenne) et en abscisse (*les articulateurs : Apex, base de la langue, dos de la langue, lèvre inférieure, mandibule, et la lèvre supérieure*). Différence significative entre le mouvement de la base de la langue et le dos de la langue comparés à ceux de la lèvre supérieure.

L'étude comparative de la mobilité des articulateurs pendant les blocages en français permet de noter des mouvements plus importants du dos de langue (4.217) suivi de la base de la langue (4.13). Pendant ces disfluences nous remarquerons que la lèvre supérieure effectue très peu de mouvement (1.621). L'analyse statique révèle que les mouvements de la base et du dos de langue sont significativement plus importants que ceux effectués par la lèvre supérieure.

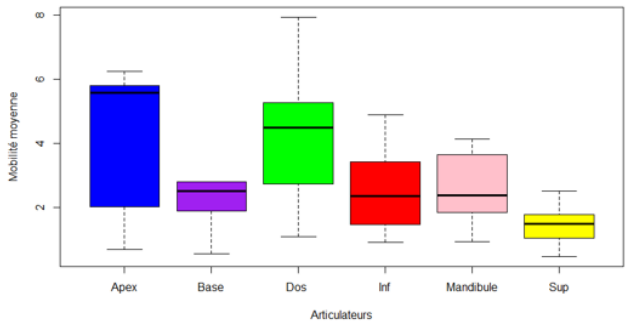


Fig.6. Mobilité des articulateurs pendant les blocages en mina. En ordonné (*Mobilité moyenne*) et en abscisse (*les articulateurs : Apex, base de la langue, dos de la langue, lèvre inférieure, mandibule, et la lèvre supérieure*). Aucune différence observée.

Pendant les blocages en mina, les mouvements du dos de la langue sont plus importants (4.253) suivis de ceux effectués par l'apex (4.151). En dernière place de la plus mobile à la moins mobile, se trouve la lèvre supérieure (1.542). Néanmoins, les différences entre l'amplitude des différents déplacements ne sont pas significatives.

3.4 Etude comparative portant sur la durée des disfluences en français et en mina

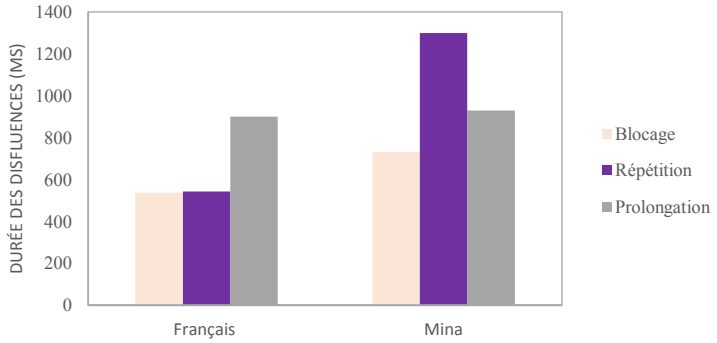


Fig.6. Durée des disfluences en français et en mina. En ordonné (*la durée des disfluences*), en abscisse (*les langues : le français et le mina*). Pas de différence significative au sein des deux langues. Mais les disfluences ont tendance à être plus longues en mina.

La figure 6 montre la durée des disfluences en fonction de la langue. En français, il est à noter que ce sont les prolongements qui sont les disfluences les plus longues (900.1 ms). La durée moyenne des répétitions et des blocages est respectivement de 542.833 et 536.888. Néanmoins, l'étude statistique ne permet pas de noter une différence significative. En mina, nous remarquons que ce sont les répétitions qui sont des disfluences les plus longues avec une durée moyenne de 1299.33 ms. Ensuite, viennent les prolongements (928.6ms) puis les blocages (730.384 ms). De même une analyse statistique ne permet pas de noter une différence significative entre les différentes valeurs.

De façon générale, il ressort que toutes les types de disfluences sembleraient plus longue en mina comparer aux durées observées en français.

4 Discussions et conclusion

Pour rappel, l'objectif de cette recherche est d'étudier les caractéristiques articulatoires et acoustiques des disfluences typiques de bégaiement dans un contexte de bilinguisme. Deux paramètres ont donc été comparés : il s'agit de l'amplitude du déplacement des articulateurs pendant les disfluences et de leur durée.

Il ressort de notre corpus que les particularités acoustiques et articulatoires diffèrent selon les deux langues.

Sur le plan articulatoire il est à noter que les mouvements pendant les disfluences sont plus importants en mina. Cette différence pourrait être due au fait de la particularité prosodique et du système tonal du mina.

De façon plus précise, au niveau des prolongations en français, les mouvements de l'apex et de la base de la langue sont significativement plus exagérés que ceux effectués par la lèvre supérieure. Alors qu'en mina, même si les valeurs sont plus élevées, les tests n'admettent pas de différences significatives.

Quant aux répétitions, aucune différence significative ne semble exister entre la mobilité des articulateurs en français, alors qu'en mina, l'apex effectue des mouvements plus importants que la lèvre supérieure.

L'étude au niveau des blocages révèle que les articulateurs ne sont pas au repos pendant ces types de disfluence. En effet dans les deux langues on observe des déplacements des articulateurs pendant les disfluences types blocages. Ces résultats corroborent ceux obtenus par Didirkova et al, 2016 dont la conclusion est que les manifestations acoustiques ne révèlent pas les caractéristiques articulatoires. En d'autres mots, les blocages, ne correspondent pas à une immobilisation des articulateurs.

Sur ce plan acoustique, on note qu'aucune différence significative n'est observée en comparant la durée des disfluences. Néanmoins, il est relevé que pour toutes les disfluences, les valeurs sont plus importantes en mina.

En conclusion, les caractéristiques articulatoires et acoustiques diffèrent d'une langue à l'autre. Néanmoins on observe très peu de différences significatives, ce qui est certainement dû au fait que cette étude a été réalisée auprès d'une seule locutrice. Il est donc important, d'augmenter le nombre de sujet afin d'analyser le plus de disfluences possibles.

Références

1. Boersma, P. Weenink, 2017. *Praat: doing Phonetics by Computer. Retrieved from*
2. Connally, E. L., Ward, D., Howell, P., & Watkins, K. E. (2014). Disrupted white matter in language and motor tracts in developmental stuttering. *Brain and language, 131, 25-35.*
3. Connally, E. L., Ward, D., Howell, P., & Watkins, K. E. (2014). Disrupted white matter in language and motor tracts in developmental stuttering. *Brain and language*, *131*, 25-35.
4. Drayna, D., & Kang, C. (2011). G*enetic approaches to understanding the causes of stuttering. Journal of neurodevelopmental disorders, 3(4), 374.*

5. Gbedahou, Dodji & Didirkova, Ivana & Hirsch, Fabrice. (2019). Caractéristiques phonétiques des disfluences typiques du bégaiement dans un contexte bilingue : le cas des locuteurs français-mina.
6. Monfrais-Pfauwadel, M. C. (2014). *Bégaiement, bégaiements : un manuel clinique et thérapeutique. Solal.*
7. Piérart, B. (2011). *Les bégaiements de l'adulte (Vol. 5).* Editions Mardaga.
8. Slim Ouni, Loïc Mangeonjean. VisArtico : visualiser les données articulatoires obtenues par un articulographe. Actes de la conférence conjointe JEP-TALN-RECITAL 2012, Jun 2012, Grenoble, France. pp.129-135
9. Van Borsel, J., Maes, E., & Foulon, S. (2001). Stuttering and bilingualism: A review. *Journal of Fluency Disorders*, *26*(3), 179-205.
10. Yairi, E., & Ambrose, N. (2013). Epidemiology of stuttering: 21st century advances. *Journal of fluency disorders, 38(2), 66-87.*

Aspects acoustiques et articulatoires des disfluences bègues dans un contexte de bilinguisme

Le cas du mina et du français

D. GBEDAHOU[1],

F. HIRSCH[1]

[1]PRAXILING, UNIVERSITÉ PAUL-VALÉRY MONTPELLIER 3,

PRAXLING

Plan

I. Généralités sur le bégaiement
II. Objectifs de l'étude
III. Méthodologie
IV. Résultats préliminaires
V. Discussion

1. Généralités sur le bégaiement

PRAXILING

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

cnrs

Définition

Trouble de l'élocution se traduisant par la survenue de disfluences altérant le rythme de la parole (Monfrais-Pfauwadel, 2014)

Trois principaux types de disfluences, pouvant être combinées :

- Répétitions
- Prolongations
- Blocages

Types et étiologie du bégaiement

Plusieurs types de bégaiement : développemental transitoire, développemental persistant, neurogène

Le *bégaiement développemental* :

- Démarre entre 2 et 7 ans
- Prévalence : entre 0,72% et 1% de la population adulte (Piérart, 2011, Yairi & Ambrose, 2013)
- Taux de rémission : 80%

Etiologie

- Origines multi-factorielles (Monfrais-Pfauwadel, 2014)
- Facteurs neurologiques (De Nil *et al.*, 2008 ; Connally *et al.*, 2014)
- Facteurs génétiques (Drayna & Kang, 2011)

PRAXILING

Phonétique du bégaiement

Sur l'anglais

Syllabes accentuées (Wingate, 1979)

Consonnes non-voisées (Jayaram, 1983, Blomgren, 2012)

Consonnes labio-dentales (Blomgren, 2012)

Consonnes constrictives (Jayaram, 1983)

Clusters consonantiques en début de mot (Wolk, 2000)

Sur le français et le slovaque

Consonnes non-voisées (Didirková, 2016)

Consonnes occlusives (Didirková, 2016)

PRAXILING

Bilinguisme et bégaiement

Fréquence du bilinguisme : environ 4000 langues parlées pour 193 pays

Bilinguisme = facteur favorisant le bégaiement ?

Manifestation du bégaiement : dans une langue ou les deux ?

Si les deux, bégaiement identique, différent ?

Le type de bilinguisme a-t-il une influence sur le bégaiement ?

Bilinguisme et bégaiement (2)

Plusieurs observations contradictoires:

Bégaiement uniquement dans la langue première (Dale, 1977) ;

Présence de disfluences pathologiques dans les deux langues, avec une sévérité différente dans ces deux langues -> *difference-hypothesis* (Shenker *et al.*, 1998; Jankelowitz & Bortz, 1996; Maruthy *et al.*, 2015);

Sévérité comparable dans les deux langues pratiquées par le sujet-> *same-hypothesis* (Van Riper, 1982; Mohammadi *et al.*, 2012)

PRAXILING

Gbedahou *et al.* (2019 A et B)

Les locuteurs bégaient moins en mina (L1) qu'en français (L2)

Manifestations différentes du bégaiement en français et en mina

- En français:
  - Voyelles antérieures > voyelles postérieures
  - Voyelles nasales > voyelles orales
  - Consonnes non-voisées > consonnes voisées
- En mina:
  - Voyelles ouvertes > voyelles fermées

II. Objectif et Hypothèse

Objectif et Hypothèse

Objectif

Vérifier si les disfluences sont plus sévères d'un point de vue articulatoire et acoustique dans une langue par rapport à l'autre

Hypothèse

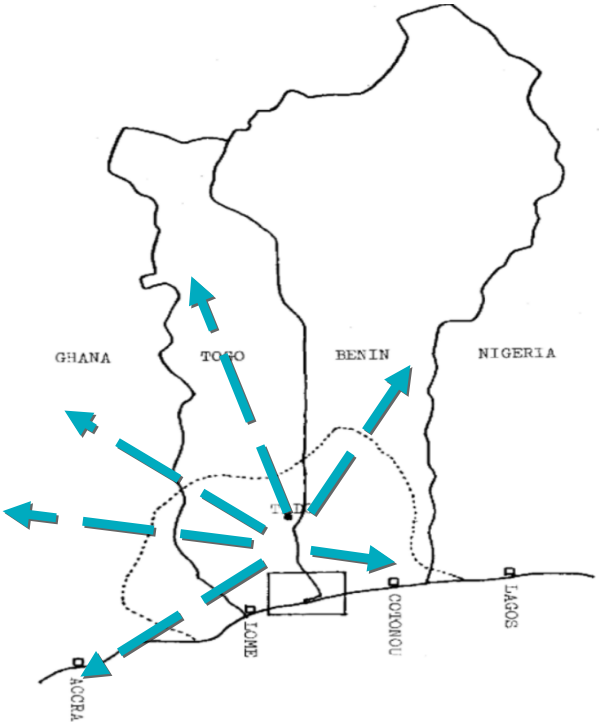
Les mouvements articulatoires seraient davantage désordonnés et anarchiques dans une langue par rapport à l'autre et leur durée ne serait pas la même en fonction de la langue.

PRAXILING

Le Mina

Expansion du Mina

Dynamisme commercial des locuteurs en particulier les « Nana-Benz »



PRAXILING

Le système phonétique du mina

	Bilabiales	Labio-dentales	Alvéolaires	Palato-alvéolaires	Rétroflexes	Palatales	Vélaires	Glottales	Labio-vélaires	
Plosives	P b		t d		ɖ	c	k g		k͡p	g͡b
Nasales	m		n		ɳ	ɲ	ŋ			
Fricatives		f v	s z				x	h		
Affriquées				d͡ʒ						
Spirantes			l						w	

Système consonantique du Mina / Adapté de Bole-Richard, 1983

PRAXILING

Système vocalique

12 voyelles (Gblem-Poidi et Kantchoa 2012)

- Sept voyelles orales : [a], [e], [ɛ], [i], [o], [ɔ], [u]
- Cinq voyelles nasales : [ã], [ɔ̃], [ɛ̃], [ĩ], [ũ]

Syllabes en mina : Syllabes V, V, VC, mais également des séquences tri-consonantiques CCCV, [klw] dans [àklwí] (bouillie locale).

Système tonal du mina

Deux tonèmes Haut et Bas

Deux réalisations modulées (Bas-Haut et Haut-Bas)

Sur le plan phonotactique, les combinaisons [V_1CV_2] comportent principalement un [a] ou un [e] comme voyelle initiale et que, dans pareille configuration, V1 est produite avec un ton majoritairement bas.

III. Méthodologie

Population et conditions d'acquisition des données

Population

Une locutrice

- Âgée de 27 ans
- L1 français, le mina à 4 ans
- Bégaiement développemental persistant

Lieu

- Laboratoire lorrain de recherche en informatique et ses applications (LORIA) Nancy

PRAXILING

Population et conditions d'acquisition des données

UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3

cnrs

Technique

Articulographe électromagnétique; (visualisation sur trois axes (X, Y et Z)

Source: Thèse Ivana Didirkova, 2016

PRAXILING

Placement des capteurs

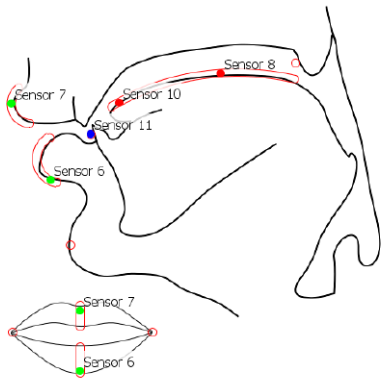
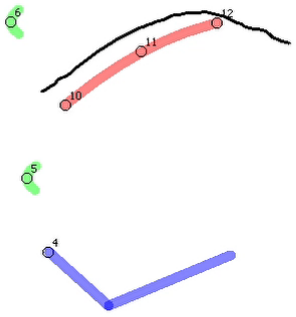
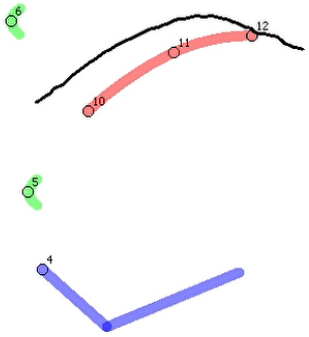


Figure 1: Placement of sensors for the mandible, the lips and the tongue.

PRAXILING

Exemples





PRAXILING

Tâches

Parole spontanée

- Questions relatives aux activités effectuées la veille + un évènement marquant dans l'actualité

Langage narratif

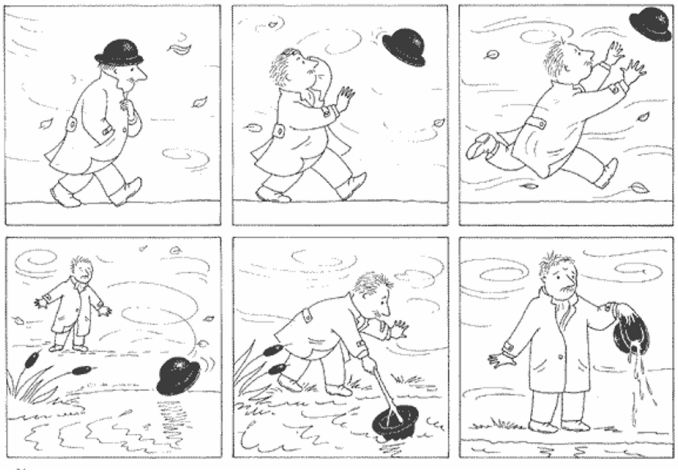
- Frog where are you? Histoire de 30 planches (Mercer Mayer)

Description d'images séquentielles

- Le sujet observe une série d'images séquentielles qu'il restitue ensuite de mémoire

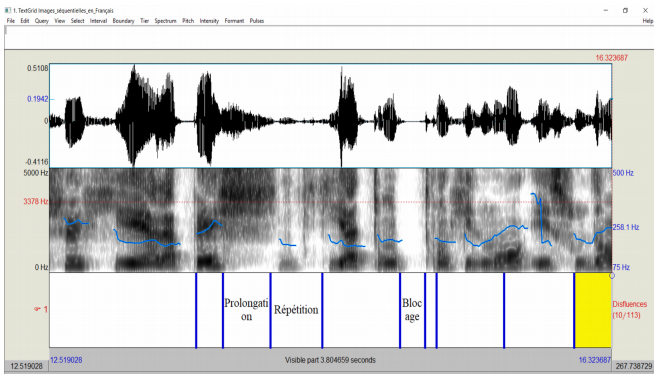
Epreuve de description d'images séquentielles

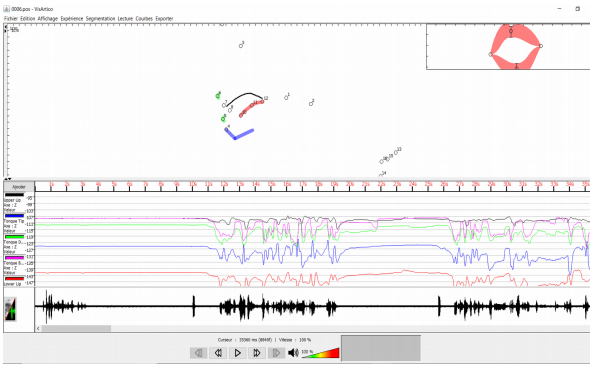
Epreuve de description d'images séquentielles



PRAXLING

Analyses





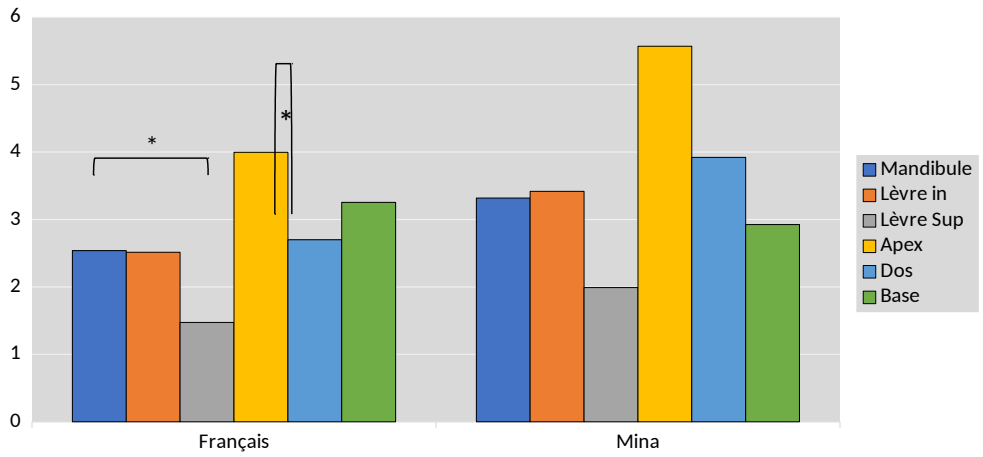
Observations

La durée des disfluences en fonction du type de disfluences et de la langue

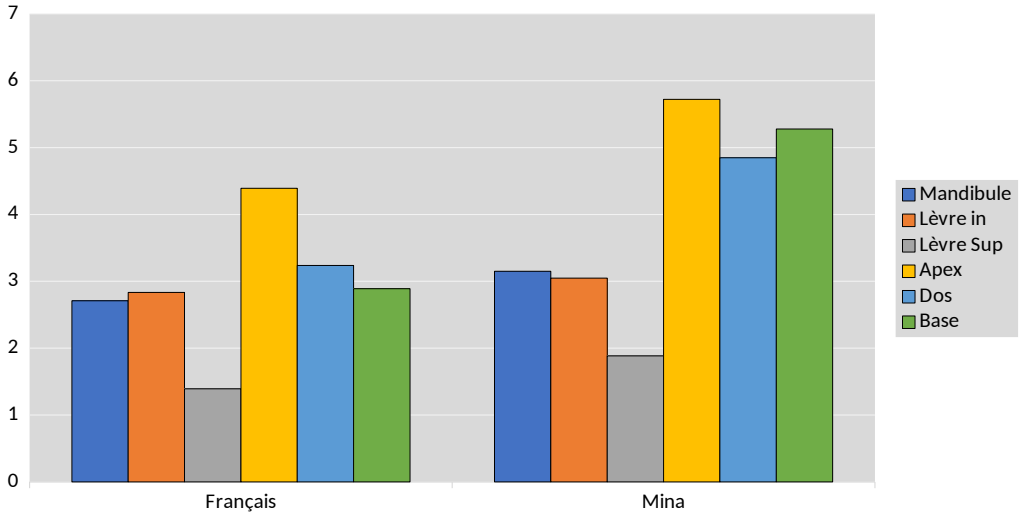
L'amplitude du mouvement des articulateurs

IV. Résultats

Mobilité des articulateurs pendant les prolongations

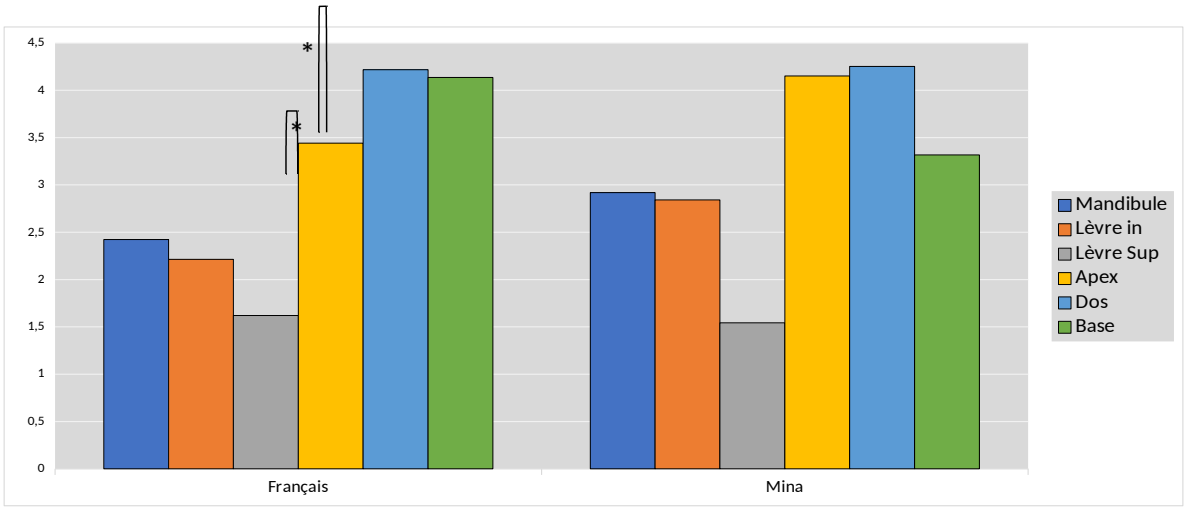


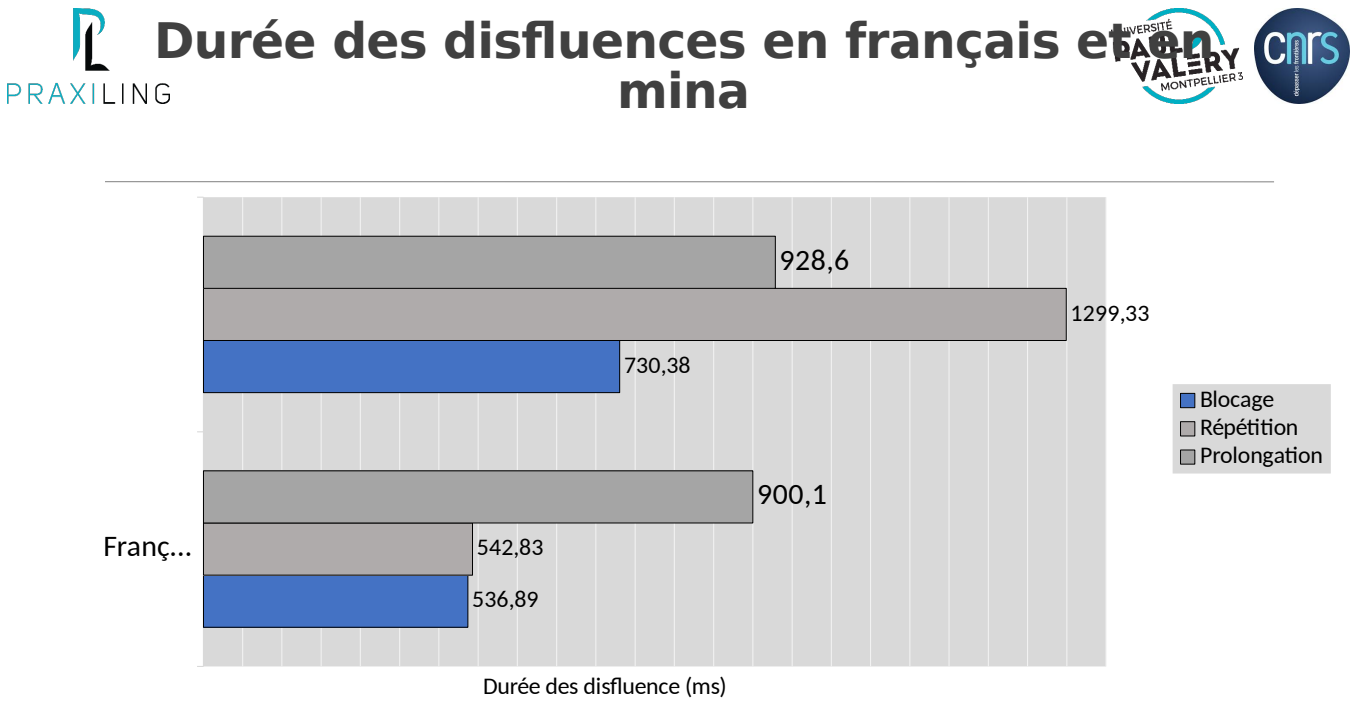
Mouvement des articulateurs pendant les répétitions



PRAXILING

Mobilité des articulateurs pendant les blocages




PRAXILING
Durée des disfluences en français en mina
928,6
1299,33
730,38
900,1
542,83
536,89
Franç...
Blocage
Répétition
Prolongation
Durée des disfluence (ms)
PRAXILING UMR 5267 - CNRS / UNIVERSITÉ PAUL VALÉRY MONTPELLIER

V. Discussion

PRAXILING

Pour résumer

- ✓ Manifestations différentes du bégaiement en français et en mina

Sur le plan articulatoire

- ✓ Mouvements plus amples des articulateurs pendant les disfluences en langue mina (tendance?)
  - ➢ Particularités prosodiques du mina?
- ✓ Blocage : présence de mouvements inaudibles (Didirkova, 2016)

Sur le plan acoustique: durée plus importantes en mina

Dans les deux langues : Apex plus mobile et lèvre supérieure moins

PRAX LING

Limites et perspectives

- Nombre de locuteurs à augmenter
- **Étudier les contextes phonétiques des disfluences**
- **Étudier les structures syllabiques disfluentes**
- Étudier le lien entre les traits tonals et l'apparition des disfluences
- Étudier d'autres niveaux linguistiques (prosodie par exemple)

Merci pour votre attention

REMERCIEMENTS : ECOLE D'ORTHOPHONIE DE LOMÉ, PROJET BENEPHIDIRE

2

Axe 2 : Numérique, cognition, apprentissages et représentations psychosociales

2.1 Le rôle du prof. dans les dispositifs numériques : une vision diachronique

François MANGENOT, LIDILEM EA609, Université Grenobles-Alpes

Depuis l'apparition des premiers logiciels d'enseignement assisté par ordinateur (EAO) s'est posée la question du rôle du prof. Dès 1980, Taylor avait théorisé le rôle de l'ordinateur dans la pédagogie (évidemment lié au rôle de l'enseignant) : ordinateur tuteur, ordinateur outil ou ordinateur « enseigné » (tutee). Dans le premier cas, la machine était censée se substituer à l'enseignant, dans le second elle devait l'aider à concevoir et à mener des activités, dans le troisième, c'était l'enfant qui était censé programmer l'ordinateur et non l'inverse (Papert, 1981). Dans tous les cas, un certain nombre de pédagogues ont estimé que c'était l'intégration du logiciel dans un scénario pédagogique qui était le plus important (Bourguignon, 2004). Puis est arrivé le web, avec la richesse de ses contenus et de ses outils de communication, qui ne rentrait bien dans aucune de ces trois catégories. Mangenot (1998) a proposé une classification des apports d'Internet à l'apprentissage des langues, prédisant que ce seraient les outils de communication à distance qui permettraient les scénarios pédagogiques les plus probants : les nombreuses pratiques de télécollaboration, très répandues en langues, sont venues appuyer ce point de vue. Enfin, le phénomène MOOC, que l'on peut considérer comme apparenté au Web social, a encore fait bouger les lignes en proposant à la fois des contenus vidéo réalisés par des spécialistes et des formes d'accompagnement par les pairs : les MOOC se sont multiplié, mais on sait peu de chose sur leur efficacité pour apprendre. Une tendance semble se faire jour, qui nous ramènerait à la première phase : l'idée, nullement prouvée, que les MOOC permettent de pratiquer la classe inversée revient à poser la question de leur intégration (en tant que SPOC, notamment) dans des dispositifs hybrides animés par les enseignants ; et c'est le degré d'intégration entre ce qui se fait à la maison et ce qui se fait en classe qui est probablement le principal garant d'une certaine efficacité pédagogique (Nissen, 2019). Le rôle du prof utilisant le numérique serait donc essentiellement celui d'un ingénieur des apprentissages, construisant des scénarios pédagogiques favorisant l'autonomie des apprenants, vieille idée datant des années 1980 (Holec, Barbot, etc.) et qui semble aujourd'hui remise au goût du jour.

2.2 MOOC : lorsque le Maître se met en scène

Didier Ozil

MOOC : lorsque le Maître se met en scène

Didier Ozil[1]

[1] Université Paul Valéry Montpellier 3, LERSEM, Route de Mende
34 199 Montpellier Cedex 5
perso@didier-ozil.fr

Résumé. Qu'ils soient MOOC, SPOC ou COOC, les « aides audiovisuelles à l'enseignement » font largement appel au langage audiovisuel. Nous nous intéresserons à celles qui mettent en scène l'enseignant lui-même, en nous appuyant principalement sur un MOOC du CNAM, *« L'intelligence artificielle pour les managers et leurs équipes »,* diffusé sur la plateforme FUN. Qui voit-on dans ces images ? Le maître, un.e acteur-trice, un.e médiateur-trice ? En filmant le « cours » du maître, la caméra réifie l'instant éphémère ; réplicable à l'infini, c'est un objet ou un ensemble d'objets, qui peut être échangé, vendu, acheté, tout ou partie. Nous appuyant sur les écrits médiologiques de Régis Debray, nous tenterons d'établir comment se manifeste la dynamique du support dans ces objets. Enfin, nous basant sur les approches de Gilbert Durand, nous définirons l'imaginaire qui se dégage - ou pas - de ces images.

Mots-clés. E-learning, MOOC, transmission, médiologie, enseignant

Abstract. Whether MOOC, SPOC or COOC, « audiovisual teaching aids » make extensive use of audiovisual language. We will be interested in those that show the teacher himself, relying mainly on a CNAM MOOC, "Artificial intelligence for managers and their teams", broadcast on the FUN platform. Who do we see in these pictures ? The teacher, an actor, a mediator? By filming the master's «course», the camera changes the ephemeral moment into an object ; infinitely copied, this object can be exchanged, sold, bought, all or part. Based on the mediology theories of Régis Debray, we will try to establish how the medium's dynamics manifest themselves in these objects. Finally, based on Gilbert Durand's approaches, we will define the "imaginary part" that emerges - or not - from these images.

Keywords. E-learning, MOOC, knowledge transfer, mediology, teacher.

Introduction

La transmission du ou des savoirs se voit proposer actuellement une panoplie d'outils nouveaux liés aux possibilités du numérique, à l'utilisation des réseaux, à l'interactivité et, de façon débutante, à l'intelligence artificielle. Dans quelle mesure l'image du « maître » est-elle modifiée, voire transformée par ces outils ? Notre objectif est d'aborder ici cette question.

On globalise souvent ces techniques et ces outils sous le vocable unique de « e-learning », ce qui donne l'apparence d'une panoplie cohérente, composée d'éléments nés en même temps. Pourtant ce n'est pas le cas et ce sera le premier point de notre analyse.

Un agglomérat d'outils divers

Loin d'être un ensemble unifié, le e-learning se présente comme un agglomérat d'outils. Simples fichiers PDF, animations graphiques, data visualization, webinaires, serious games, etc., tous ces éléments possèdent leur propre histoire et propre code. Les MOOC, qui désignent les « *massive online open courses* » se combinent avec les outils cités. Loin d'être des « objets » unifiés, ils sont également des ensembles d'unités ; par exemple une succession de films. D'une certaine façon, les MOOC sont des « conteneurs ».

Nous avons évoqué des « codes » différents pour ces différents outils. Les MOOC, par exemple, utilisent des films ou des séquences filmées qui ne sont autres que des prolongements de la « vidéo » apparue dans les années 1950, puis constamment améliorée et simplifiée jusqu'à aujourd'hui. Les codes d'écriture de ces films ou séquences filmées ont été élaborés au cours de ces années et l'on pourrait même remonter vers des expériences antérieures, par exemple les expériences de télévision scolaire [1].

Comme le note Mathieu Cisel, le point fort des MOOC ne tient pas dans une « révolution technique », ni même « pédagogique » [2]. Le « tsunami MOOC », pour reprendre son expression, est lié au double OO de l'acronyme : Open Online. Plus que l'objet lui-même, c'est sa mise à disposition sur internet qui créé le phénomène[3], l'acronyme désigne d'ailleurs indifféremment les plateformes qui accueillent les cours et les cours eux-mêmes. La métonymie est ici parlante.
Si l'on tient la date de 2001, comme date fondatrice du phénomène, date de création de l'Open Coursware du MIT, on mesure le chemin parcouru jusqu'à aujourd'hui en regardant les chiffres de 2016 [4] :

1 En 1963, la télévision diffuse 10 heures de programme par semaine. Voir le sujet INA, n° de notice : I09345614. Titre : Télévision scolaire. Émission du 3 novembre 1963. Participants : P. Chilotti, directeur de l'Institut Pédagogique National (IPN). H. Dieuzeide, responsable de la radio télévision scolaire à l'Institut Pédagogique National. Durée : 7 min 25 s.

2 Cf le blog de Mathieu Cisel *La révolution MOOC* et plus particulièrement l'article *MOOC : comment tout a commencé*. Du même auteur, on peut également consulter l'article : *Chronique des MOOC*, revue Sticef, Sciences et Technologies de l´Information et de la Communication pour l´Éducation et la Formation, site sticef.org.

3 Il n'est pas inutile ici de rappeler que l'enseignement à distance possède également son histoire : le CNED est créé en 1939 pour tenter de remédier à la désorganisation de l'enseignement due à la guerre puis à l'occupation.

4 Lien vers l'article de la revue EdSurge, entreprise de presse dédiée à l'éducation, initialement financée par la Graham Holdings Company - https://www.edsurge.com/news/2016-12-29-monetization-over-massiveness-breaking-down-moocs-by-the-numbers-in-2016

Entreprise	Nombre d'inscrits
Coursera	23 millions
edX	10 millions
XuetangX	6 millions
FutureLearn	5.3 millions
Udacity	4 millions

Cadre de notre analyse

Notre analyse va se centrer sur un MOOC, qui nous servira d'étude de cas.
Dans les formations sans présentiel, le maître n'est présent qu'au travers d'une image filmée ; la représentation que l'apprenant peut avoir de lui ne peut se faire qu'à travers une « représentation » imagée.
Le double sens du mot « image » est ici capital: il y a l'image « numérique », qui est un ensemble de pixels, produit par une machine, et l'image au sens de la « représentation » que se fabrique l'apprenant.
Dans quelle mesure cette représentation est-elle importante dans le processus de transmission ?

Éducation et autorité

Pour répondre à cette question, nous nous appuierons sur une phrase de Durkheim.

> *« Ce qui fait l'autorité dont se colore si aisément la parole du prêtre, c'est la haute idée qu'il a de sa mission ; car il parle au nom d'un dieu dont il se croit, dont il se sent plus proche que la foule des profanes. Le maître laïc peut et doit avoir quelque chose de ce sentiment. Lui aussi, il est l'organe d'une grande personne morale qui le dépasse: c'est la société. De même que le prêtre est l'interprète de son dieu, lui, il est l'interprète des grandes idées morales de son temps et de son pays.»* [5]

Cette phrase nous l'interprétons ainsi : la qualité de la transmission du ou des savoirs est liée à une représentation que l'apprenant se fait du maître, laquelle représentation dépasse le niveau de diplôme du maître, son savoir-faire pédagogique, son expérience, sa personne même, etc., mais renvoie à un imaginaire situé en dehors de lui, en-dehors de l'Institution d'enseignement elle-même, tout en étant à l'intérieur du groupe social.

5 Durkheim E. *Éducation et sociologie*, 2016, PUF. [1922]. Page 68.

Il ne s'agit pas, bien entendu, d'instituer l'enseignant « prêtre » mais qu'il soit considéré comme tel. Et ce « comme » nous renvoie ici à la pensée analogique : tout en n'étant pas identique à celui auquel je suis comparé, j'en possède les qualités.
Notre projet est donc d'essayer de voir si cette autorité morale peut avoir, aujourd'hui, un équivalent, autrement dit « à quoi doit ressembler l'enseignant d'aujourd'hui » pour être auréolé de cette autorité ?

Que nous disent images et sons ?

La plateforme FUN s'impose comme plateforme de premier plan en France : créée en 2012, elle annonce deux millions d'inscriptions en 2016 [6].
Au sein de la plateforme FUN, un MOOC attire l'attention : « *Du manager agile au leader designer* » qui arrive en tête du classement des années 2013-2017, tous classements confondus avec 113 357 inscriptions.
Pour autant ce n'est pas ce MOOC, un peu ancien, sur lequel s'est porté notre choix, mais un autre, réalisé par la même équipe, diffusé selon les mêmes modalités et mis en ligne en 2018 : *« L'intelligence artificielle pour les managers et leurs équipes »* - production CNAM, diffusion FUN.

Vue d'ensemble du MOOC

Ce MOOC propose un déroulé pédagogique sur quatre semaines.

Introduction
Direct
Information attestation vs certification
S1 – Comprendre ce qu'est l'IA - Test hebdomadaire 1 - Forum thématique
S2 – Observer des cas d'usage - Test hebdomadaire 2 - Forum thématique
S3 - Tirer des leçons - Test hebdomadaire 2 - Forum thématique
S4 – Développer de nouvelles compétences - Test hebdomadaire 2 - Forum thématique
Test final
Certificat
À la suite de cours

6 Source site FUN.

Particularités du MOOC

Premier constat, ce MOOC comporte beaucoup de films, plus de 50. Avec une amplitude de 2 minutes à 20 minutes, ils ont une durée moyenne de 7 min 30 s. L'ensemble représente environ 7 heures de film. Ce qui est un « métrage » important pour une formation de quatre semaines.
L'autre point important, c'est qu'il y a essentiellement des films. Le reste des items sont quelques pages de graphismes avec parfois des renvois, par exemple pour des compléments d'information.
Le sujet de ce MOOC est une présentation de l'intelligence artificielle au travers d'un grand nombre de cas d'entreprises qui utilisent cet ensemble de techniques, ainsi que quelques enseignants spécialisés. Le point fort de ce cours est d'aller à la rencontre de ces experts et de les interroger sur leur lieu d'exercice : Berkeley, Paris, Singapour, Shenzhen, Hong Kong...

La professeure qui présente ce MOOC et qui l'a conçu, Mme Dejoux, est présente dans tous les films.
Un premier type de situation nous montre l'enseignante face à la caméra, souvent devant un fond blanc, avec quelques textes ou graphismes en incrustation dans l'autre partie de l'écran ; posture que nous qualifierons de « magistrales ».
Mais pour la plupart, ces films sont des interviews d'une ou deux personnes, interviews qui sont assurées par la professeure. Il y a 73 personnes différentes qui sont interviewées dans ce MOOC. Dans six cas, l'intervieweuse est face à deux personnes et dans cette configuration, elle se place toujours entre les deux personnes à interviewer.

Nombre de minutes de l'enseignante seule (posture magistrale)	Interviews (posture d'intervieweuse)
1 h 22 min	5 h 10 min + 40 s (démo tutoriel) + 5 min 19 s (démonstration d'une machine)

Cette omniprésence assure une grande continuité dans le MOOC, et également une incarnation maximale, une « présence ». Sont ainsi évités les « PowerPoint » muets ou accompagnés d'une voix sans corps, d'une voix dite « off ».

Cette continuité est également due à la posture, au type de « jeu » utilisé par l'enseignante qui déploie une grande énergie : elle travaille sans notes, avec le regard fixé dans la caméra, et ainsi, garde un contact constant avec l'apprenant.

7 Le terme de « film » est ici synonyme de « séquences ». Aucune de ces séquences n'est assez autonome pour être qualifiée de « film », ce terme désignant plus habituellement un objet cohérent. Néanmoins, parler ici de film a du sens puisque, nous le verrons plus tard, chacune des séquences est détachable de l'ensemble.

Analyse du MOOC

Pour entrer dans le détail des films, nous utiliserons un type d'approche critique que l'on emploie habituellement pour analyser des films de cinéma. Nous nous sommes appuyés sur le type d'analyse d'Anne Goliot-Lété 8, enseignante à Paris Diderot. Cette approche se propose de décomposer un film en différentes composantes : l'image, le son, le montage ; à l'intérieur de l'image le cadrage, les mouvements de la caméra, à l'intérieur du son, la musique, etc. Il s'agit d'un type d'analyse qui distingue nettement une phase de déconstruction et détaille un film ou une séquence comme on le ferait d'un produit chimique, en analysant un par un ses composants. Ce type d'analyse possède, bien entendu, ses limites. D'une part, il faut accepter l'idée que la somme des parties d'un film est égale au tout, ce qui peut être contredit dans certains cas. D'autre part, ce type d'analyse ne prend pas en compte le regard du spectateur. Mais dans le cas qui nous occupe, cette prise en compte est complexe, elle ne peut se faire qu'au travers du forum du MOOC, donc de façon partielle et biaisée.

Analyse de l'image

Ce qui apparaît en premier lieu, sur le plan global, c'est le budget de production réduit de l'ensemble. Cela se traduit par des problèmes d'image, des réglages qui se font en direct, « à vue », par exemple quelques rectifications de lumière, de cadrage, de zooms un peu hésitants. Les interviews sont réalisées à une seule caméra sauf quatre d'entre elles qui sont réalisées à deux caméras. Un certain nombre d'interviews sont réalisées avec un téléphone portable, que l'on aperçoit furtivement dans le reflet d'une vitrine. La constatation principale que nous faisons de ces interviews monocaméra, ce qui saute d'emblée aux yeux, c'est la frontalité des cadrages.
Cette frontalité apparaît logique dans les séquences que nous qualifierons de magistrales, où l'enseignante est filmée de façon classique devant une sorte de tableau blanc, un dispositif très proche de celui de la salle de cours. Mais elle se retrouve également pendant les interviews, où la disposition des interviewés et de l'intervieweuse est inhabituelle par rapport à ce qui se pratique : un-e intervieweur-se qui prend la place du spectateur et donc que l'on retrouve de dos pendant l'interview.

Au niveau des regards, un jeu de l'enseignante est particulièrement interpellant. Pendant les interviews, l'intervieweuse regarde l'interviewé, mais régulièrement elle se retourne vers les apprenants comme pour les prendre à témoin et, dans le même temps, rétablir une liaison par le regard avec eux. Voilà qui transgresse également une autre norme : soit on regarde la caméra et on s'adresse au spectateur, soit on prend symboliquement sa place, mais le spectateur ne peut alors se regarder lui-même.

8 Nos références se feront via un ouvrage de cet auteur (cosigné par Francis Vanoye) chez Nathan Université en 2001 [parution initiale : 1992] et notamment les premières pages de l'ouvrage qui posent la méthode d'analyse.

9 *« Décomposer en ses éléments constituants (...) que l'on ne perçoit pas à l'oeil nu »* [ouvrage cité, page 9].

Nous qualifierons volontiers ces regards-caméras de « percussions » pour souligner à la fois leur caractère inédit et la rupture, l'accroc qu'ils apportent dans le tissu de l'interview.

Il faut analyser les interviews à deux caméras pour retrouver une norme extrêmement établie qui est celle du champ contre champ. Cette figure très majoritaire dans le langage audiovisuel propose une alternance assez mécanique, montrant alternativement celui ou celle qui pose les questions et celui ou celle qui y répond, et, parfois, comme une sorte de dérogation, celui ou celle qui écoute.

Son

Il n'y a aucun travail sur le son, adjonction de musique ou de sons additionnels (bruitages, etc.). Là encore, c'est le matériau brut qui est fourni et cette « brutalité du réel » se traduit souvent par la disparité des niveaux sonores.

Le montage

Les séquences d'interviews sont réduites au recueil de témoignages : il n'y a pas d'images d'illustrations, par exemple un interviewé que l'on voit en train de travailler ; pas de fioritures, pas de temps morts. Dans le même temps, le dispositif de filmage apparaît parfois comme transparent au spectateur : on laisse des hésitations, on voit l'enseignante qui vient couper la caméra à la fin d'une interview, on perçoit la télécommande qu'elle utilise pour enchaîner les différents blocs du MOOC.
L'autre point marquant, sur les interviews mono caméra, est l'absence de montage au sein de l'interview : sauf dans de très rares exceptions, nous sommes ici dans un « tourné-monté » qui privilégie l'énergie et la continuité d'une prise unique.

Conclusion de l'analyse

L'ensemble se présente donc comme une alternance de séquences qui structurent des moments pris sur le terrain, sur le vif, qui gardent de leurs captations une sorte « d'énergie du réel ». L'absence de tout aspect « lisse », absence qui n'est pas du tout recherchée, nous en sommes convaincus, authentifie les discours, à la fois le discours de l'enseignant et celui des interviewés. En abandonnant une valeur-spectacle, le MOOC se retrouve crédité d'une valeur-témoignage indubitable.

Nous sommes face à une forme qui n'est ni lisse ni léchée ou aseptisée, et qui, par conséquent, laisse l'énergie de la fabrication intacte, loin d'un objet normé.

Statut de l'enseignante à l'image et problème du genre

Qui voit-on dans ces films ? Lorsqu'il s'agit de séquences magistrales, il n'y a pas de doute : l'image véhicule le point de vue de la salle de cours, avec un cadrage assez large, tout en permettant de bien visualiser l'expressivité du visage de l'enseignante. Dans ce cas, l'enseignante est bien une experte, celle qui sait et qui transmet.

Dans les interviews, le statut de l'enseignante disparaît ou se fond derrière celui d'une intervieweuse, d'une journaliste, pourquoi pas d'une médiatrice, voire d'une actrice. Plus exactement, le statut de l'enseignante reste le même, mais le corps qui est filmé semble se démultiplier dans diverses tâches.

Pour vérifier la place que prend ce corps filmé dans l'image, nous proposons un test double que nous intitulons « rapport corps-décor » et qui se présente sous la forme de deux chiffres. Le premier chiffre indique le rapport entre le pourcentage de surface qu'occupe l'enseignement par rapport à l'ensemble de l'image, le second caractérise l'arrière-plan de la séquence, selon un classement que nous avons effectué :

Niveau	Caractérisation	Exemple
5	Fond blanc ou noir, décor naturel sans signification particulière, fonds flous	
4	Fond de type 5, mais comportant des graphismes, des animations, de la data visualization.	
3	Environnement signifiant par rapport au sujet, mais fixe, nature morte	fond d'affiches, de posters, etc.
2	Environnement signifiant et en mouvement	arrière-plan de bureaux avec va-et-vient de personnes, etc.
1	Fond qui apporte réellement un plus de signification au sujet traité	

Sur cinq séquences représentatives analysées, les résultats sont assez homogènes.

Type de séquence	Rapport corps-décor	Type de fond
Introduction magistrale devant un téléviseur faisant office de tableau	10 %	4
Introduction d'une interview	11,25 %	4
Interview deux personnes	17,90 %	1
Interview trois personnes	11,74 %	3
	Moyenne (sur dix séquences) 15,58 %	Moyenne 3,31 (itw) 4 (magist.)

Ces chiffres expriment que l'enseignante est très peu filmée en gros plan [10]. Que ce soit lors des séquences magistrales ou pendant les interviews, sa présence à l'écran est similaire.
Le cadrage fait en sorte qu'elle conserve le même statut, qu'elle interroge les industriels (position d'intervieweuse qui renvoie, très souvent dans les différents films, à la position du candide), ou qu'elle établisse des synthèses de ces interviews (position d'expert).

Le problème du genre
Ce MOOC n'appartient pas au genre télévisuel, non plus au genre documentaire.

Dans le genre télévisuel, un journaliste assure le lien avec le téléspectateur, mais il s'efface le plus souvent au moment des interviews dans une technique de champ / contre champ.

Le MOOC n'appartient pas non plus au genre documentaire qui, même si la forme documentaire est très souple, accorde une importance à l'image que l'on ne retrouve pas dans cette succession de séquences, très resserrées sur leur sujet.

Ce que l'on peut en conclure c'est que le MOOC représente un genre particulier avec ses propres codes : présentation très frontale de l'image, regards-caméra que la professeure adresse à son public-apprenants, discours continu et de forme identique que l'enseignante soit en position d'expert ou en position d'intervieweuse, utilisation d'une gamme de plans très resserrée, avec des tailles de plans presque identiques dans l'ensemble du MOOC (rapport corps-décor quasi identique).

L'ensemble est cohérent et il tend, à nos yeux, vers un but : laisser constamment à l'enseignant son statut de professeur, de passeur de savoirs qu'il possède en présentiel.

Dans une interview qu'elle donne au journal *Le Monde*, Cécile Dejoux revient, indirectement, sur ce problème du genre, tel qu'elle le perçoit, en deux citations très éclairantes :

> *il faut raconter des histoires, être à la fois théorique et pratique, être multi-outils : vidéo, liens internet, donner des quiz...»*
> *Quand vous avez potentiellement 30 000 personnes qui peuvent réagir sans filtre, c'est de la formation télé-réalité ! 11*

10En comparaison, un MOOC de Coursera, avec un type de cadrage très attendu (du type de celui du journal télévisé) donne des résultats autour de 30 pour cent.

11Journal « Le Monde », *La « Mooc-mania » gagne la France*, article de Nathalie Brafman, publié le 28 mars 2014 à 15h44 - Mis à jour le 04 septembre 2014.

Ni reportage télé, ni documentaire, le MOOC relève plus, pour cette enseignante, d'un phénomène de masse. La référence à la télé-réalité induit également une autre idée : le passage, très actuel, du film unitaire à la série. Pour cette enseignante, le MOOC ne possède pas la continuité d'un film mais se remet en jeu chaque semaine face à un auditoire qu'elle doit conquérir.

La réification du « cours »

À partir du moment où une caméra filme l'intervention d'un enseignant, le statut du cours change : il se transforme en objet. Dès lors, l'objet-cours peut être dupliqué, cédé, vendu et transformé. Dispensé devant un public de 50 apprenants, il peut être vu par 200, 2 000, etc. ; il entre dans une logique de reproduction mécanique. Cette transformation en objet change-t-elle le statut de l'enseignant, la nature de son enseignement ?
Autrement dit, les différents aspects techniques que nous avons analysés le définissent-ils dans sa totalité ? Le support-objet modifie-t-il la transmission ? Celle-ci peut-elle disparaître pour ne laisser qu'un objet « purement technique » ?

Dans un premier temps, nous renvoyons volontiers à cette boutade de Régis Debray qui indique que 2+3 = 5 est une équation qui reste vraie quel que soit le support qui la véhicule [12].

Au-delà, revenons sur la distinction que cet auteur établit entre « transmettre » et « communiquer » :

> *Une transmission est une communication optimisée par un corps, individuel ou collectif.* [13]
> *De même y a-t-il des machines à communiquer, mais non à transmettre (…). La transmission ajoute à l'outil matériel de la communication un organigramme, en doublant le support technique par une personne morale. (…) La transmission est charge, mission, obligation : culture.* [14]

Ainsi donc, le MOOC garde le sens d'une transmission lorsqu'il s'appuie sur une institution qui le commandite, institution qui est ici représentée par l'enseignante. C'est ce qu'elle représente - l'organisation qui se cache derrière les différentes séquences – qui donne son sens à l'ensemble.

Pour autant le support garde-t-il une dynamique propre ? Le MOOC distribué par la plateforme FUN repose sur un triangle composé d'un contenu, d'un « dispositif technique et d'une Institution » 15.

12Debray, R. (2000) *Introduction à la médiologie.* Presses Universitaires de France. Page 33.
13 Debray, R. *ouvrage cité*, page 4.
14 Debray, R. (1997) *Transmettre.* Odile Jacob. Page 20.
15 Debray, R. Ouvrage cité, tableau page 15.

Dès lors que le MOOC est copié, remonté, amputé, redistribué, son caractère d'agglomérat reprend le dessus et la dynamique de cohésion disparaît. La « valeur-transmission » disparaît pour laisser place à une « valeur-communication ». Son centre de gravité, pour reprendre les termes de Régis Debray, passe de « pour mémoire » à « pour usage ».

Nous voyons, dans cette perte d'identité possible, l'une des limites du MOOC. Cette limite est aussi celle au-delà de laquelle la transmission du savoir, ainsi que les enseignants qui en sont les vecteurs, sont entraînés dans une parodie simplifiée, voire contre-productive, de « communication ». Cette limite n'est d'ailleurs pas liée aux MOOC, mais bien à l'image filmée, produite par une caméra, laquelle ne filme que des corps et décors et non des essences.

Le rôle cognitif de l'image

Interrogerons l'image en elle-même. Dans un ouvrage de 1994, Gilbert Durand note la prédominance d'un courant qui affirme le rôle « cognitif » de l'image- c'est-à-dire « produisant une connaissance ». Pour Durand, ce courant éclate depuis près d'un demi-siècle et il le nomme « révolution vidéo ». Il oppose ce courant à une « production obsessionnelle […] contingentée dans le domaine du *distraire* » [16].

Cette opposition entre deux statuts de l'image renvoie au fait que certains media, comme la télévision, peuvent être suspectés de favoriser massivement des programmes de distraction. Au-delà de ça, cette idée traduit que l'image pose encore problème, de par sa nature polysémique (Roland Barthes, 1964 [17]). Cette complexité semble souvent mal maîtrisable et conduit à une utilisation normée, simplifiée, de l'image dont on peine à utiliser l'ensemble des ressources qu'elle offre.

C'est ce que l'on observe dans ce MOOC. Il suffit de regarder les différents films pour voir que l'image est ici, essentiellement, le filmage d'une parole. Bien entendu l'image apporte un effet de réalité : les personnes qui sont interviewées l'ont réellement été. Mais l'image n'a servi qu'à enregistrer leur dire et non à montrer leur faire. La caméra est ici, à la limite, réduite à un rôle d'enregistreur de surveillance.

De fait, le sujet même, l'intelligence artificielle, n'est pas de nature très visuelle. On évoque par exemple des chatbots, donc des robots programmés pour répondre par chat à des utilisateurs et, par définition, ces chatbots, qui sont des logiciels, ne sont pas "filmables" en l'état. Pourtant, une séquence nous montre bien comment l'intelligence artificielle analyse une page Linkedin, sous une forme qui évoque le tutoriel. De même, une autre séquence nous montre une démonstration d'une machine « lego », mise en place pour vulgariser l'application de l'IA à l'usine de demain.

16 Durand, G. (1994). *L'imaginaire, essai sur les sciences et la philosophie de l'image*. Paris: Hatier. Page 20 - 22.

17 Barthes, R. (1964) *Rhétorique de l'image*. In: Communications, 4, 1964. Recherches sémiologiques. pp. 40-51.

L'image à laquelle se réfère G. Durand est l'image du réel, remplie de sens, celle qui « a donc le statut de symbole, type même de la pensée indirecte, où un signifiant avouable renvoie à un signifié obscur » [18].
Partant de cette fonction de l'image, que nous qualifierions volontiers de « magique », Durand évoque ensuite une « pédagogie positiviste ». L'image nous permettrait de sortir d'un raisonnement binaire « vrai -faux » en apportant « un réel voilé », là où la logique aristotélicienne exige « clarté et distinction ».

Ce débat semble dépasser nos constats et pourtant il les prolonge. Si on analyse son texte, l'image qu'évoque G. Durand est une image porteuse d'un acte artistique, et non le simple enregistrement d'une machine, aussi perfectionnée soit-elle.
Le MOOC peut être le lieu où l'on donne une chance à l'image, dans toute sa complexité, mais cette mise en place doit être assumée par une vision qui ne peut être celle d'un simple opérateur.
En d'autres termes, si la « mise en sens » est le fait de l'enseignant, qui se trouve ici tout à son rôle, si la « mise en forme » est celle d'un opérateur technique, sans que cette dénomination ne comporte rien de péjoratif, une troisième approche est, pour nous, nécessaire, celle d'une « mise en scène », qui permette à l'image audiovisuelle de dépasser un statut d'image de caméra de surveillance, qui ré investisse les richesses et les subtilités du langage audiovisuel.
À ce prix, l'image pourrait être porteuse d'un univers pédagogique autre et concourir à *« un appel à l'activité de l'élève plutôt qu'à sa docilité »*, pour reprendre la formule de J. Piaget.

Conclusion

L'interview, déjà citée, du journal *Le Monde,* nous servira de conclusion.
Dans cette interview, la professeure Cécile Dejoux tire un bilan de son travail sur les différents MOOC.

> « Lorsque vous avez fait un MOOCs, vous ne faites plus jamais un cours devant un auditoire comme avant. C'est une vraie aventure pédagogique.»
> « 50% de la réussite de ce cours, c'est la plateforme, 50% ce sont les réseaux sociaux ».

Le terme d'aventure pédagogique montre bien, s'il en était besoin, que le e-learning et l'ouverture indispensable vers les réseaux sociaux, notée par Mme Cécile Dejoux dans cet article, appellent des enseignants empreints d'une certaine «élasticité», terme que nous empruntons à Jacques Perriault [19].

18 Durand, G. (1994). *Ouvrage cité,* page 23.
19 Cité par Jacques Wallet, *À l'heure de la société mondialisée du savoir, peut-on supprimer les enseignants ?* Hermès, la revue. CNRS Editions. 2006/2 n° 45, pages 91 à 98.

Le maître de demain (ou d'aujourd'hui!), devra-t-il s'inspirer, pour repenser son ministère, de l'analyse de Régis Debray et de son découpage de l'Histoire en trois « périodes » : Logosphère (période de l'écriture), Graphospère (période de l'imprimerie), vidéosphère (période de l'audiovisuel) ? Nous relèverons, dans le tableau qu'il nous propose, les mythes d'identification associé à chaque âge :

Ages	Logosphère	Graphosphère	Vidéosphère
Mythes identification	Le Saint	Le Héros	La Star

Ce mythe d'identification renouvelle d'une certaine façon, la représentation de l'enseignant en prêtre, proposée par Durkheim ; elle semble nous indiquer que les techniques de *e-learning* nous font entrer dans une ère où la *connivence* des enseignants avec l'image sera *a minima*, requise.

Annexes

Structure du MOOC

Introduction

- 0-1 - Présentation du MOOC
  - 0-1-1 - film de 7 min 5 s. Présentation de Cécile Dejoux, interview de Joël de Rosnay, expert
  - Texte sur le MOOC
  - CV de la professeure
  - 0-1-2 - film de 20 s. Présentation de Quentin, community manager des MOOC
  - texte d'incitation à se présenter avec discussion associée
- Questionnaire de départ
  - texte et questionnaire en ligne
- Livres du MOOC
  - biblio
- De l'intelligence artificielle au collaboratif
  - page d'incitation à l'inscription (facultative) à l'autre MOOC : « Du manager au leader : devenir agile et collaboratif »
- Chaîne Youtube
  - page d'incitation à l'inscription (facultative) sur la chaîne Youtube avec lien direct
- Newsletter
  - page d'incitation à l'inscription (facultative) à la newsletter du Learning Lab
- réseaux sociaux du MOOC
  - page d'inscription (facultative)
- Avertissements
  - avertissements liés à la navigation hors du site FUN, notamment avec les vidéos « interactives », diffusées à partir d'une autre plateforme.

Direct

- Vos rendez vous en directement
  - dates de connexion possibles pour questions directes au professeur
- Participez au directement
  - page de liens pour webconférence
- Enregistrements à voir en différé
  - Deux enregistrements de webconférence

Information attestation vs. Certification

- Attestation gratuite
  - Texte d'explication
- Certification payante
  - Texte d'explication

S1 – Comprendre ce qu'est l'IA

- Questionnaire sur l'IA
  - Questionnaire en ligne

- S1-1 – Qu'est ce que l'IA ?
  - Film de 11 min 47 s. Présentation du sujet avec fond d'image illustration graphiques et photos. Interview de Denis Florean, IBM : définition de l'IA.
- S1-2 – Pourquoi aujourd'hui ?
  - Film de 10 min 44 s. Présentation CD, interview d'Arnaud Lariche (Ernst & Young), désannonce, data viz sur maturité et coût, volume des investissements, valorisation boursière, interview de Nathanaël Ackerman (Hub France IA), désannonce, dataviz sur les différents domaines, renvoi sur une vidéo youtube de chirurgie par IA, interview de Bertrand Braunsweig, INRIA) : véhicule autonome, désannonce et conclusion
- S1-3 - Quels sont les domaines d'application
  - Film de 4 min 38 s. Graphismes sur les différents domaines d'applications. Interview de Olivier Ezratty, consultant. Désannonce, graphismes et photos sur les domaines d'intégration.
- S1- 4 – Débats et enjeux
  - Film de 16 min 01 s. Introduction sur trois débats : peur vis a vis du travail, peur vis à vis d'objets dans le corps humain, peur par la création de robots. Interview de Sophie Pene, universitaire Paris Descartes. Schéma sur l'exemple du chatbot. Interview d'Alexandre Cadain, fondateur d'Anima. Débat n° 2 : bioconservateurs contre transhumanismes. Pas d'interview pour ce débat et enchaînement sur la troisième peur : les humanoïdes. Interview de Dan Fisher, General manager de l'entreprise Infiniti. L'IA destructeur du manager ? Interview de Christophe Alnette, Netgem International à Singapour, désannonce puis sujet IA créateur de métiers ? Interview Sophie Pène. Désannonce et conclusion du film.
- Test hebdomadaire 1
- Forum thématique

S2 – Observer des cas d'usage

- S2-1 – Cas d'usage
  - Film S2-1, 12 min 48 s. Introduction sur les différents cas d'usage, deux cas ensuite sur téléphone portable : Malkoff Médéric et EDF. Annonce des différentes interview à venir.
    Film S2-1-1, 9 min 33s. Interview : Stéphane Lalee, Facebook, Singapour.
    Film S2-1-2, 6'51''. Interview Leo Chen et Chuanxiong Chzzen, Université chinoise.
    Film S2-1-3, 8'08''. Interview David Marchesseau, SAP, Singapour, Innovation Center.
    Film S2-1-4, 5'16''. Interview Evrard Dauger, manager industrie EY.
    Film S2-1-5, 7'48''. Interview William Eldin, start up XXII. Sujet sur l'analyse de flux vidéo en temps réel. Exemple Casino. Exemple sécurité en aéroport-gare.
    Film S2-1-6, 3'22''. Interview M. Bo An, NTU, Singapour. Exemple Ali Baba.

Film S2-1-7, 6’51’’. Interview Hugo Garcia Cotte, fondateur Cyphen, Shenzhen.
Film S2-1-8, 5’40’’. Interview de Stephane Waller, fondateur Bleexo. Logiciel de mesure de l’implication.
Film S2-1-9, 2’12’’. Interview de Eric Dupuy.
Film S2-1-10, 6’10’’. Interview de Christoiphe Montagnon et Mailis Ruol, Randstadt. Interview dans un lab d’innovation, entretien sur un chatbot.
Film S2-1-11, 8’41’’. Interview de Djamil Kemal, fondateur Goshaba.
Film S2-1-12, 6’04’’. Interview de Matthieu Dejardins, Next User. Analyse de la voix, analyse des comptes sociaux, par exemple pour une marque de shampoing, capture de candidats potentiels pour l’entreprise en analysant les compte sociaux.
Film S2-1-13, 5’33’’. Interview de Nicolas Bertin
Film S2-1-14, 7’19’’. Interview de Robert Chu
Film S2-1-15, 2’22’’. Interview de Jeremy Harroch, Quantmetry
Film S2-1-16, 9’53’’. Interview de Florence Estra
Film S2-1-17, 6’01’’. Interview de Emmanuelle Joseph

- S2-2 – Les cas en version mobile sur Teach on Mars
  - Page de renvoi vers l’application sur téléphone mobile
- S2-3 – Le cas Malakoff-Médéric-Humanis
  - Film S2-3-1, 9’07’’. Présentation de la stratégie, la méthode et quelques cas. Interview d’Emmanuel Copin, DRH de MMH. Interview de Hakim Mamouri, sur le sujet : en quoi le business model de MMH évolue ? Retour sur Emmanuel Copin.
  - Film S2-3-2, 16’10’’. présentation de la méthodologie. Trois axes. Interview de Gauthier Lalande, manager AI et block chain. Interview Anne Sophie Godon, Innovation chez MMH ? Interview de Marc Fargeas, études stratégies MMH.
  - Film S2-3-3. 20’58’’. Les projets MMH. Présentation des différents projets par CD. Interview de Gauthier Lalande, projet un : détection de la fraude optique. Partie deux : les compétences du manager. Projet deux : Interview de Hakim Mamouri sur la valeur ajoutée de l’IA. Projet trois : les robots payeurs. Interview de Christophe Cretel sur les robots payeurs pour les retraites. Interview d’Odile Prauca. Manager Cicas, relevé de carrière pour retraités sur tablettes.
  - Film S2-3-4, 7’09’’. Présentation par CD. Quatre chapitres ; Partie 1 : GPEC et conduite du changement RH. Interview de Marie-Laure Godfroy-Legrand. Part 2 : outils et métiers. Interview de Henrique Dos Santos. Part 3 : valeur ajoutée dans le business model, poursuite de l’interview de Henrique Dos Santos. Part 4 : qu’en pensent les collaborateurs ? Poursuite de l’Interview.
- S2-4 – Le cas EDF
  - Film S2-4-1, 6’58’’. Introduction du cas : gérer la complexité. Interview de Aude Vinzerich, responsable pôle IA EDF, explication des objectifs.
  - Film S2-4-2, 7’30’’. Méthodologie : présentation par CD. Interview de Aude Vinzerich, responsable pôle IA EDF.

    - Film S2-4-3, 7’56’’. Projet 1 : le chatbot numérique. Interview de Julie Amiot, juriste et Idriss El Asry, expert IA.
    - Film S2-4-4, 5’55’’. Projet 2 : Gecko. Présentation CD. Gecko : rapport d’incident pour les agents EDF. Interview d’Adèle Desoyer, expert IA, Simon Devaradja, ingénieur IA. Dexcription du projet, valeur ajoutée,
    - Film S2-4-5, 0’35’’. Conclusions du cas EDF par CD : l’IA n’est pas là pour remplacer des personnes.
  - Test hebdomadaire 2
  - Forum thématique

S3 -tirer des leçons

- S3-1 – Remplacer : quelles sont les tâches que l’IA va réaliser à la place du manager ?
    - Film s3-1, 9’45’’. Évaluer. Évaluation des systèmes d’IA. Présentation des tâches de l’IA : remplacer (le manager), Interview de Arnaud Vianney de Microsoft. Classer des documents, Interview de Christopher Muffat, entreprise Dathena, Singapour. Prendre des rendez-vous, Interview de Praveen Velu, entreprise Evie, Singapour. Répondre à des questions, Interview de Thomas Jestin, KRDS, Singapour.
- S3-2 – Assister : quelles sont les tâches que le manger fait avec l’IA ?
    - Film S3-2, 5’42’’. Assister le manager. Interview Arnaud Vianney de Microsft, explication du tri de documents ou recherche de documents pour apporter des compléments d’information. Faire de la curation et bulle filtrante. Interview de Benoît Raphael, éleveur de robots au sein de Flint. L’analyse de sentiments, avec l’analyse de texte. Interview de Denis Florean, IBM. La gestion de projets, pas d’interview.
- S3-3 – Augmenter : quelles sont les nouvelles tâches que le manager va faire grâce à l’IA ?
    - Film S3-3, 12’15’’. Augmenter. Présentation CD : « réinventer le métier de manager ». Exemple 1, traduction et analyse de documents. Interview de Arnaud Vianney de Microsoft. Exemple 2 : l’innovation, aide à la découverte de nouvelles idées. Interview de Christophe Lecante, pdg de TKM. Exemple 3 : créativité. Créativité musicale par exemple avec Sony Music. Interview François Pachet, Spotify. Exemple 4 : Optimisation des lieux de travail. Interview de Jean Marc Deromedi, fondateur de Aprim, Singapour puis Interview de Sébastien Bréal, fondateur de Locarise, Tokyo. Conclusion avec tableau récapitulatif.
- S3-4 – Identifier quels sont les facteurs clés de succès ?
    - Film S3-4-1, 8’33’’. Présentation CD : savoir mener un projet IA. Présentation des différentes interviews suivantes.
    - Film S3-4-2, 8’11’’. Interview de Jeremy Harroch.
    - Film S3-4-3, 2’38’’. Interview de Juliana Lim.
    - Film S3-4-4, 2’58’’. Interview de Martin Plouy.
    - Film S3-4-5, 6’59’’. Interview de David et Philippe Remaud.
    - Film S3-4-6, 8’33’’. Interview de David Remaud et Bertrand Mabille.
    - Film S3-4-7, 2’46’’. Interview de Jeremy Harroch. Quantmetry, aide au commercial itinérant.

    - Film S3-4-8, 7’19’’. Interview de Robert Chu. Copie du fichier présent en S2-1-14.
- S3-5 – Les leçons à tirer
    - Film S3-5-1, 7’43’’. Résumé récapitulatif de CD : l’IA sujet mondial, sujet sociétal.
    - Film S3-5-2, 8’33’’. Interview de Julien Bence
    - Film S3-5-3, 6’51’’. Interview de Hugo Garcia Cotte. Copie du fichier présent en S2-1-7.
    - Film S3-5-4, 6’29’’. Interview de Feng Yuan
- Test hebdomadaire 3
- Forum thématique

S4 – Développer de nouvelles compétences

- Questionnaire sur l’IC (intelligence créative
- S4-1 – Créer des systèmes informants
    - Film S4-1, 3’40’’. Présentation CD sur les compétences d’interaction hommes-IA. Interview d’Oliver Guérin, Artik Consulting.
- S4-2 – Comment éduquer les IA et apprendre des IA
    - Film S4-2, 7’07’’, Entraîner une IA, Interview de Benoît Raphael, éleveur de robots et CEO de Flint.media. Agréger les intelligences, Interview de François Taddei, CRI. Apprendre des IA, avoir une culture de la donnée, Interview d’Avner Bar-Hen, enseignant au CNAM.
- S4-3 – Contrôler les IA
    - Film S4-3, 6’28’’. Trois approches : contrôler les biais, Interview d’Antoine Borde, chercheur chez Facebook, mettre des systèmes d’alerte Interview de William Dab, directeur de chaire au Cnam, savoir débrancher, laisser l’humain dans la boucle, Interview de Manuel Zacklad, professeur au Cnam.
- S4-4 – Problématiser et avoir l’esprit critique
    - Film S4-4, 5’07’’. Questionner, sortir du cadre, repenser le monde du travail. Interview de Fabrice Fisher, fondateur de Blu. Interview de Ross Garvie, Infini Nissan, Hong Kong : ce qu’on attend d’un manager c’est qu’il suscite. Sur la problématisation, Interview de Praveen Velu, co fondateur d’Evie.
- S4-5 – Acculturer à l’IA
    - Film S4-5, 5’31’’. Expliquer ? Expérimenter, accompagner l’intégration, repenser les métiers. Interview de Eric Thain sur le changement de mindset, head of digital Lewis. Interview de David Daoud, elearn2grow, Hong Kong sur la formation et la nécessité de se former
- S4-6 – Penser la complémentarité
    - Film S4-6, 5’30’’. Interview sur l’intelligence collective de Joël de Rosnay : charisme, vision, valeurs, écoute, la confiance : les valeurs de Joel de Rosnay. Interview de François Taddei, directeur du CRI, sur l’entreprise apprenante. Conclusion assurée par CD avec tableau texte récapitulatif.
- S4-7 – Imaginer le futur du travail

  - Film S4-7-1, 15’27’’. Présentation du problème par CD qui annonce les intervenants. Puis interview de Gary Bolles, directeur de la chaire futur du travail, singularity university, Californie. Interview de Viviane Ming, professeur à Berkeley, Soccos Lab.
  - Film S4-7-2, 4’12’’. Reprise de l’Interview de Gary Bolles, incluse dans le film S4-7-1.
  - Film S4-7-3, 8’28’’. Reprise de l’Interview de Vivienne Ming, incluse dans le film S4-7-1.
- Test hebdomadaire 4
- Forum thématique

Test final
Certificat
A la suite de cours

- Questionnaire de fin
- Pour aller plus loin

2.3 L'intégration, l'usage et l'apport de la compétence technologique dans l'enseignement-apprentissage des élèves du primaire : quels jalons et quels enjeux

2.4 Enquête sur la littératie numérique des étudiants chinois pour l'apprentissage du français

Yigong Guo

Enquête sur la littératie numérique des étudiants chinois pour l'apprentissage du français

Yigong Guo[1]

[1] ENS de Lyon, ICAR UMR 5191, 15 Parvis René Descartes BP 7000, 69342 Lyon Cedex 07, France
yigong.guo@outlook.com

Résumé. La littératie numérique occupe une place importante dans les compétences nécessaires aux citoyens du XXI[e] siècle. Cet article a pour objectif de présenter les résultats d'une enquête destinée à faire un état des lieux de la littératie numérique dont disposent les étudiants chinois pour l'apprentissage du français. Les résultats montrent que ce public a accès à une diversité d'équipements informatiques et est capable de s'en servir ; mais il peine à dépasser certains usages ordinaires, bien que la plupart des étudiants adoptent une attitude positive vis-à-vis de l'intégration des TIC dans l'apprentissage du français.

Mots-clés. didactique des langues, littératie numérique, pratique numérique

Abstract. Digital literacy is acknowledged as one of the key competences which should be acquired by all citizens in the 21st century. The purpose of this article is to present the results of a survey which aims to assess the digital literacy of China's French-major students. The results show that the students have access to a diversity of digital technologies and are able to use them in different learning situations; but they struggle to go beyond certain ordinary uses, although most of them adopt a positive attitude towards the integration of ICT into French language learning.

Keywords. language learning and teaching, digital literacy, digital usage

1 Introduction

À l'ère de la culture numérique, de nouvelles exigences apparaissent dans le domaine de l'éducation et de l'emploi. La littératie numérique occupe ainsi une place importante dans les compétences nécessaires aux citoyens du XXI[e] siècle [4]. Cet article a pour objectif de présenter les résultats d'une enquête destinée à faire un état des lieux de la littératie numérique dont disposent les étudiants chinois dans leur apprentissage du français.

Nous commencerons par une brève définition de la notion de littératie numérique, avant de préciser son rapport avec l'enseignement-apprentissage des langues. Nous aborderons ensuite la méthodologie mise en œuvre dans le cadre de ce travail, et présenterons les résultats d'analyse qui permettent de mettre en lumière les pratiques

numériques et les attitudes des étudiants vis-à-vis de l'intégration des TIC dans leur apprentissage.

2 Littératie numérique pour l'enseignement-apprentissage des langues

Devant le rythme soutenu du changement et l'introduction continue de nouvelles technologies, les citoyens du XXI^{e} siècle doivent tenir à jour leurs compétences numériques qui leur permettront de s'adapter au changement [6]. Au cours des vingt dernières années, plusieurs termes ont été utilisés pour désigner ces compétences : computer literacy, ICT literacy, media literacy, information literacy et digital literacy [2]. Depuis les cinq dernières années, un consensus émerge sur le terme de littératie numérique, plus englobant, significatif et générique que les autres [3].

La littératie numérique implique l'acquisition d'« un ensemble de compétences, de consciences et d'attitudes pour utiliser efficacement les TIC dans diverses situations »[1] [8] :

- « Compétence » se réfère à l'aptitude opérationnelle à mobiliser des savoir-faire essentiels pour utiliser de manière stratégique des outils numériques [1] ;
- « Conscience » renvoie à la dimension critique et éthique impliquant une réflexion sur « le comportement approprié lié à une situation et à un environnement numérique spécifiques » [10] ;
- « Attitude » porte sur l'intérêt à s'engager dans des communautés et des réseaux sociaux à des fins culturelles, sociales et/ou professionnelles [6].

[1] the awareness, attitude and ability of individuals to appropriately use digital tools [...] in the context of specific life situations. Traduction de l'auteur.

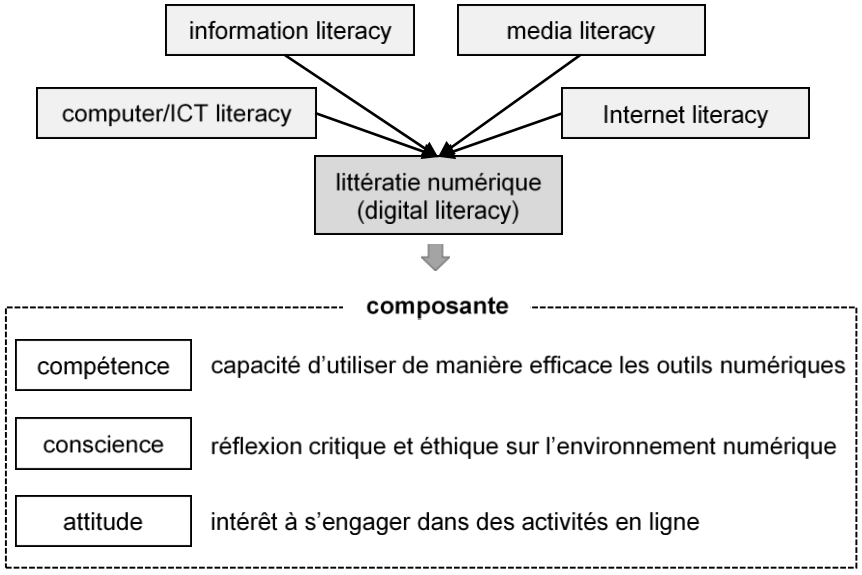


Fig. 1. Définition de la littératie numérique.

Dans le domaine de l'enseignement-apprentissage des langues, l'acquisition de la littératie numérique aide les apprenants « à construire des environnements personnels d'apprentissage multi-supports et à devenir des usagers des langues autonomes et compétents » [5]. En effet, les apprenants ayant acquis une littératie numérique sont censés être capables de mobiliser de manière efficace les ressources numériques pour mettre à profit l'opportunité d'apprentissage des langues.

3 Question de recherche

Dans cet article, nous nous intéressons à la littératie numérique des étudiants chinois spécialisés en français langue étrangère (FLE), en nous demandant dans quelle mesure ces étudiants sont compétents pour intégrer de manière efficace les outils numériques pour leur apprentissage. L'objectif est donc de contribuer à une meilleure compréhension des pratiques numériques des étudiants, de leur conscience et de leur attitude par rapport aux activités d'apprentissage avec le numérique en classe et en dehors de la classe.

4 Méthodologie

Pour répondre à cette question, une enquête a été menée auprès d'un échantillon d'étudiants spécialisés en FLE à l'Université de Zhejiang Gongshang.

4.1 Questionnaire

Un questionnaire a été distribué aux étudiants en début de deuxième et de troisième année universitaire (N=42), c'est-à-dire ayant atteint le niveau A2/B1[2]. L'enquête par questionnaire avait pour objectif de connaître leur degré d'équipement en matériels informatiques, leurs pratiques numériques pour l'apprentissage du français, et leurs attitudes relatives aux activités d'apprentissage avec les TIC.

4.2 Journal de bord

Les étudiants en troisième année (N=20) ont été invités à tenir un journal de bord, dans lequel ils racontaient leurs pratiques et leurs ressentis par rapport aux activités d'apprentissage avec les TIC tout au long du semestre. Ces données ont été ensuite complétées par des entretiens semi-directifs et par des notes d'observation des activités numériques en classe.

4.3 Traitement de données

Notre méthode de traitement relève d'une démarche descriptive et fait appel à la triangulation des différents types de données recueillies. Il s'agit : 1° d'appliquer les statistiques descriptives et inférentielles sur les résultats du questionnaire, et 2° d'appliquer l'analyse de contenu catégorielle, en catégorisant et en comptant les occurrences relatives aux pratiques numériques, à la conscience et aux attitudes identifiées dans les journaux de bord.

5 Résultats

5.1 Possession / accessibilité aux matériels numériques

Comme le montre la figure 2, le public interrogé enregistre des taux d'équipement élevés, ce qui tend à montrer que ce dernier dispose d'une diversité de matériels numériques dont ils peuvent se servir pour l'apprentissage des langues.

[2] Étant donné que le programme de l'enseignement du français en Chine n'est pas défini selon le Cadre européen commun de référence pour les langues (CECRL), le niveau de français des participants est estimé selon une étude empirique qui vise à situer le niveau linguistique des étudiants chinois au sein des 6 niveaux du CECRL [11].

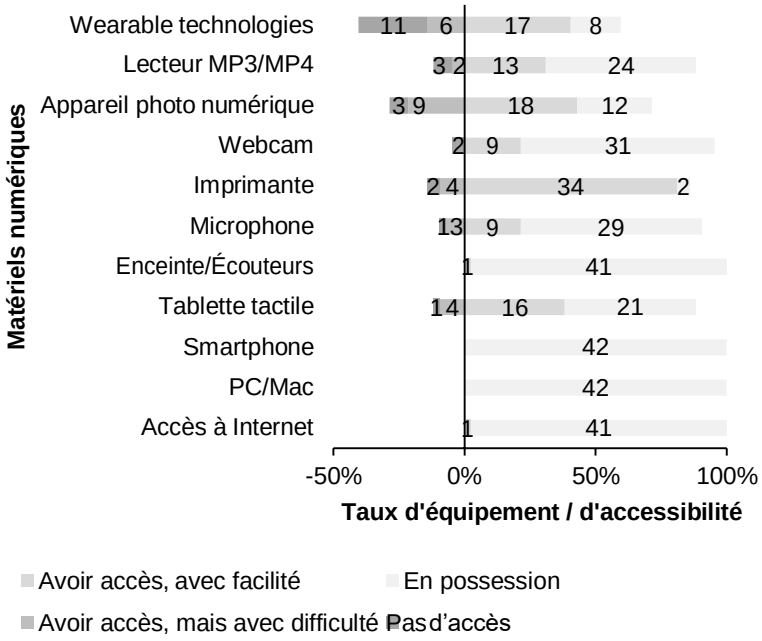


Fig. 2. Taux de possession / d'accessibilité en matériels numériques.

L'importance du taux d'équipement pourrait s'expliquer par le fait que la diversification des produits sur le marché permet aujourd'hui à une majorité de la population d'être en mesure d'acheter des équipements informatiques à moindre coût. Comme le fait remarquer Gerbault [7], la perspective actuelle nous amène à concevoir une nouvelle version de la fracture numérique, qui n'existe plus entre ceux qui ont et ceux qui n'ont pas les matériels, mais entre ceux qui savent et ceux qui ne savent pas s'en servir de manière efficace.

5.2 Pratiques numériques pour l'apprentissage du français

Les participants à l'enquête par questionnaire ont été invités à indiquer sur une échelle de cinq points allant de 1 (jamais) à 5 (très souvent), la fréquence d'utilisation des TIC pour réaliser les différentes activités d'apprentissage du français.

Table 1. Activités d'apprentissage numériques les plus fréquentes.

Activités	Moyenne	Écart-type
Consulter un e-dictionnaire	4,881	,633
Utiliser un logiciel de traduction	4,405	1,037
Utiliser un moteur de recherche pour trouver des ressources d'apprentissage	4,405	,798
Préparer un exposé avec Powerpoint	4,405	,734
Télécharger des ressources d'apprentissage à partir d'un fond de ressources	4,095	1,100
Ensemble des activités d'apprentissage	**2,858**	**,436**

α = ,856

Le tableau 1 illustre les cinq activités d'apprentissage numériques les plus fréquentes. Il s'avère que toutes les activités mentionnées n'impliquent pas les apprenants dans les démarches communicatives et interactionnelles.

Un test-t d'échantillons appariés a été appliqué, et montre une différence significative de la moyenne des scores entre les activités communicatives (M=3,132, SD=,479) et celles qui sont non communicatives (M=2,066, SD=,615), p=,000<,05 ; d=1,677 (voir le tableau 2).

Table 2. Comparaison de la fréquence des activités d'apprentissage numériques sans communication et avec communication.

Nature de l'activité	N	Moyenne	Écart-type	t	p
Activités sans communication	42	3,132	,479	10,869	,000**
Activités avec communication		2,066	,615		

** Le résultat est significatif au niveau ,01.

Les résultats obtenus permettent de constater que l'empan des usages numériques est restreint : les TIC sont souvent utilisés par les étudiants comme outils de recherche et de consultation des ressources, mais l'aspect communicationnel est rarement exploité.

5.3 Attitude pour l'intégration des TIC dans l'apprentissage du français

Dans le questionnaire, il était demandé aux répondants d'attribuer un score allant de 1 (pas du tout d'accord) à 5 (tout à fait d'accord) sur une série de questions (ex. « avec les TIC, je peux trouver facilement des documents authentiques pour l'apprentissage du français »), afin d'auto-évaluer leur attitude vis-à-vis de l'intégration des TIC pour l'apprentissage du français.

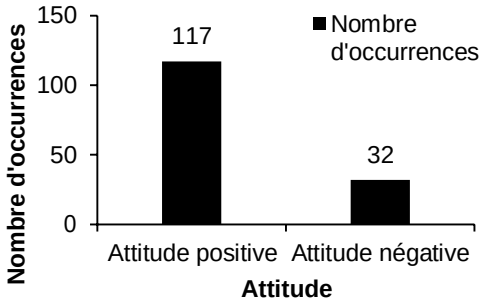


Fig. 3. Résultats de l'analyse catégorielle des journaux de bord selon les attitudes des étudiants vis-à-vis de l'intégration des TIC dans l'apprentissage du français.

D'après les résultats obtenus, la moyenne des scores est supérieure à la médiane de l'échelle de Likert (M=3,791, SD=,495), ce qui signifie que les étudiants interrogés adoptent une attitude positive et considèrent que l'intégration du numérique contribue à leur apprentissage du français. Ce résultat a été ensuite confirmé par l'analyse des journaux de bord, dans lesquels nous avons codé au total 149 occurrences qui manifestent les différentes attitudes concernant l'utilisation des TIC dont 78,523% renvoient à un aspect positif (voir la figure 3).

5.4 Conscience critique et éthique liée à l'utilisation des TIC

Il s'est agi de coder dans les journaux de bord, les marques linguistiques montrant que les étudiants étaient conscients de la dimension critique et éthique liée aux pratiques numériques. Voici un exemple qui montre que l'étudiant est capable d'adopter un regard critique sur la qualité des ressources repérées sur Internet :

> Dans une époque de l'information, nous pouvons transmettre et partager rapidement des informations. […] Mais comme dans le domaine des médias, [où sont véhiculées de fausses informations], les ressources d'apprentissage sur Internet ne présentent pas toujours la même qualité.

Nous avons enregistré au total 37 occurrences qui recouvrent le cadre critique et éthique de la littératie numérique. Ce résultat manifeste que le public possède des aptitudes à la pensée critique pour gérer le contenu et adopter des comportements appropriés dans des environnements numériques.

5 Conclusion et perspective

Pour conclure, l'enquête permet de mettre en lumière la littératie numérique dont disposent les étudiants chinois spécialisés en FLE :

- Le public bénéficie d'un niveau d'équipement informatique élevé qui permet de prendre la mesure de l'intégration des TIC dans les études universitaires ;
- Pourtant, l'empan de leurs usages en apprentissage du français est restreint, et se limite à des pratiques ordinaires, bien que la plupart des étudiants adoptent une attitude positive vis-à-vis de l'intégration du numérique ;
- Ce public est capable de mener une réflexion critique sur le contenu du web et est conscient de la dimension éthique liée à l'environnement numérique.

Dès l'apparition d'Internet et du web social, de nombreux chercheurs ont souligné les apports, dans le domaine des langues, de la communication médiée par les technologies [9]. À partir des résultats issus de notre étude, il semble que les dimensions communicatives sont rarement exploitées par les étudiants chinois. Il s'agit donc maintenant de souligner l'enjeu et la nécessité, pour l'enseignant, d'aider et d'encourager les apprenants à explorer les potentiels interactionnels du web pour leur apprentissage du français.

Références

1. Ala-Mutka, K. (2011). Mapping Digital Competence: Towards a Conceptual Understanding. Seville: JRC-IPTS.
2. Bawden, D. (2001). Information and digital literacies: A review of concepts. Journal of Documentation, 2(57), 218-259.
3. Brotcorne, P. (2014). Des compétences numériques à la littératie numérique : Quels enjeux derrière l'évolution de ces concepts. Bruxelles : Association pour une Fondation Travail- Université.
4. Carretero, S., Vuorikari, R., & Punie, Y. (2017). DigComp 2.1: The Digital Competence Framework for Citizens with Eight Proficiency Levels and Examples of Use. Luxembourg: European Union.
5. Centre européen pour les langues vivantes. (2016). Vers une littératie numérique pour l'enseignement et l'apprentissage des langues. Consulté 18 février 2017, à l'adresse http://www.ecml.at/ECML-Programme/Programme2016-2019/Digitalliteracy/tabid/1797/language/fr-FR/Default.aspx
6. Commission européenne. (2007). Compétences clés pour l'éducation et la formation tout au long de la vie : Un cadre de référence européen. Luxembourg : Union européenne.
7. Gerbault, J. (2012). Littératie numérique. Les nouvelles dimensions de l'écrit au 21ième siècle. Les Cahiers de l'Acedle, 9(2), 109-128.
8. Martin, A., & Grudziecki, J. (2006). DigEuLit: Concepts and tools for digital literacy development. Innovation in Teaching and Learning in Information and Computer Sciences, 5(4), 249-267.
9. Ollivier, C. (2014). Vers une approche interactionnelle en didactique des langues et une extension du domaine de la tâche – Les atouts du web 2.0. Alsic. Apprentissage des Langues et Systèmes d'Information et de Communication, (17).

10. Ollivier, C., & Projet e-lang. (2018). Littératie numérique et approche socio-interactionnelle pour l'enseignement-apprentissage des langues. Strasbourg, Graz : Éditions du Conseil de l'Europe.
11. Xu, Y. (2011). Test de repérage du niveau linguistique des étudiants chinois dans la perspective de l'adaptation du CECR au milieu universitaire de Chine. Synergie Chine, (6), 167-176.

2.5 Des doctorant-e-s SHS face à l'innovation numérique : enjeux techniques, politique, pédagogiques.

Camille Roelens, Marie Valorge, Aurélie Doelrasad

Des doctorant·e·s SHS face à l'innovation numérique Enjeux techniques, politique, pédagogiques

Camille Roelens[1], Marie Valorge[2], Aurélie Doelrasad[2]

[1] INSPE Académie de Lille Hauts-de France, laboratoire RECIFES (EA 4520), 365bis Rue Jules Guesde, 59650 Villeneuve-d'Ascq, France
camille.roelens@espe-lnf.fr ; roelens.camillejean@orange.fr
[2] Université Lumière Lyon 2, 86 rue Pasteur, 69007, Lyon, France
m.valorge@univ-lyon2.fr
aurelie.doelrasad@univ-lyon2.fr

Résumé. Notre article propose une réflexion théorique à propos des compétences acquises par les doctorant·e·s en sciences humaines et sociales grâce à l'expérience de la recherche dans un monde numérique. C'est notamment en construisant le cadre théorique de leur thèse que les jeunes chercheurs en SHS sont confrontés au processus de consultation, sélection, usage ou non-usage des travaux de recherche par le biais du monde numérique. Notre réflexion montre que ce processus est source d'acquisition de nouvelles compétences comme l'adaptabilité, la créativité et l'intelligence collective.

Mots-clés. Innovation numérique, Doctorants, Compétences, Choix individuels, Travail collaboratif

Abstract. Our article offers a theoretical reflection on the skills acquired by doctoral candidates in the humanities and social sciences through research experience in a digital world. It is in particular by building the theoretical framework of their thesis that young SHS researchers are confronted with the process of consultation, selection, use or non-use of research works through the digital world. Our reflection shows this process is a source of the acquisition of new skills such as adaptability, creativity and collective intelligence.

Keywords. Digital innovation, Doctoral students, Skills, Individual choices, Collaborative work

1 Introduction

Notre article s'inscrit dans le cadre des actions menées par le laboratoire junior Usages et Critiques de l'Innovation Pédagogique. Comme son nom l'indique, le laboratoire junior UCIP s'intéresse à la question de l'innovation pédagogique, en proposant un regard critique à partir des différents ancrages théoriques investis par les doctorant·e·s du laboratoire ECP. Dans ce sens, nous proposerons de considérer l'expérience doctorale à l'aune de *l'innovation numérique*, en faisant converger des cadres théoriques et des ancrages fondamentaux pluriels. Notre contribution est de nature essentiellement épistémologique : il s'agit pour chacun d'entre nous d'adopter une posture réflexive, et de dialoguer sur cette base. Comme l'écrit Marcel Gauchet

[1] il existe des matrices problématiques qui s'imposent dans des époques données (le numérique étant assurément l'une de ces matrices à l'heure actuelle), et la confrontation constructive des approches singulières vis-à-vis de cette manière partagée est des plus féconde. Ainsi, nous partirons chacun successivement des apports de notre propre thèse, pour faire converger nos interrogations vers la même tension problématique du « choix » fait par les doctorant·e·s au cours de la thèse. Nos apports sont articulés en relation au syntagme de « monde numérique », qui « met l'accent sur la dimension anthropologique, plutôt que strictement technique, des environnements virtuels, ainsi que sur des aspects fondamentaux de la relation humaine selon des fonctions qui ne sont, ni strictement rationnelles, ni directement utiles : informationnels, communicationnels (solidarités par exemple), cognitifs, culturels, mais aussi imaginaires, poétiques, spirituels » [2].

Dans une première partie, nous tenterons d'identifier et d'analyser les enjeux du développement professionnel des doctorant·e·s en sciences humaines et sociales, où le choix du thème de recherche puis celui de son cadre théorique sont fondamentaux, et qui se déploie parmi un foisonnement de ressources propre au *monde numérique*. À partir d'un état des lieux succinct des recherches dans le domaine des technologies pour l'éducation, nous présenterons dans une deuxième partie une « déconnexion »
[3] entre l'enthousiasme accordé de façon dominante au numérique pour la diffusion des connaissances, et des résultats scientifiques (relatifs à l'impact des technologies éducatives sur le développement des compétences des apprenant·e·s des enseignements primaire, secondaire et supérieur) plutôt réservés sur les capacités des technologies à soutenir les apprentissages. Nous relèverons aussi parmi cette même littérature, des résultats qui attribuent de potentielles autres capacités aux technologiques numériques pour le développement des sujets. Dans la dernière partie de notre communication, nous rapporterons ces premières réflexions à l'expérience de la formation doctorale, à partir de la démonstration que c'est véritablement une manière nouvelle d'être-au-monde et d'être en société qu'il s'agit de penser au sein du monde numérique.

2 La Formation Doctorale en SHS à l'Aune de l'Innovation Numérique

Notre réflexion est centrée sur le développement professionnel des doctorant·e·s en sciences humaines et sociales. Nous choisissons de distinguer les doctorant·e·s en sciences humaines des doctorant·e·s de « sciences exactes », à l'instar de Gérard [4]. Cette distinction tient notamment, selon cette auteure, au fait que les doctorant·e·s en sciences exactes se spécialisent dans une technique au cours de leur doctorat, et que cette spécialisation ne conditionne pas leur carrière future. A contrario, en sciences humaines, les jeunes chercheur·e·s se spécialisent sur un thème, et cela conditionne leur carrière future. Dans son ouvrage, le doctorat est considéré comme un « rite de passage » au cours duquel la·le doctorant·e acquiert des compétences grâce à l'expérience qu'il vit, et c'est dans cette continuité que s'inscrit notre réflexion : l'expérience du doctorat est un processus pendant lequel la·le jeune chercheur·e se développe professionnellement. Nous entendons le développement professionnel

comme « un processus d'internalisation qui conduit l'acteur à faire l'expérience d'une acculturation professionnelle, un processus d'externalisation qui se manifeste à travers un savoir agir et interagir en contexte » [5].

Dans le cas du·de la jeune chercheur·e, le savoir agir que nous souhaitons aborder est sa capacité à établir le cadre théorique de sa recherche et à se positionner vis-à-vis de la multiplicité des ressources disponibles, composée par la littérature scientifique et les communications avec les pairs. Cette tâche consiste à explorer les travaux de recherche, notamment aux moyens de dispositifs sociotechniques, pour en prendre connaissance, en extraire les concepts clés, et choisir de les mobiliser, ou non, dans le cadre de la thèse. C'est bien par l'expérience de la recherche documentaire et de la communication avec les pairs que la·le doctorant·e parvient à définir son cadre théorique et parvient à se développer professionnellement en tant que chercheur·e. Les doctorant·e·s développent-ils des compétences professionnelles en lien avec le *monde numérique* dans lequel ils puisent ? Si oui, quelles sont ces compétences ?

3 Innovation Numérique et Innovation Pédagogique

Flichy [6] a relevé qu'Internet avait été imaginé, créé et utilisé par les scientifiques, pour la fondation d'une « communauté scientifique idéale », sans hiérarchie, où la compétition serait moindre et stimulerait la collaboration, pour l'émergence d'une « intelligence collective ». Dans un article de 2019, Fluckiger a analysé plusieurs documents et rapports publiés ces dix dernières années par le ministère de l'éducation français, avec des résultats de la recherche en sciences de l'éducation en regard. Alors que les sources officielles associent *innovation technologique* et *innovation pédagogique*, dans la continuité de l'« imaginaire utopique d'Internet et de ses pionniers », selon les résultats scientifiques, la diffusion des technologies numériques ne coïncide pas toujours avec leur appropriation [7] par les act·eur·ice·s des terrains de l'enseignement et de l'apprentissage. Toujours selon Fluckiger [3], si les discours des décideurs ne trouvent pas d'écho dans les résultats des chercheur·e·s, c'est parce que les regards portés sur le numérique divergent : la vision technocentrée des politiques, des commerçant·e·s et des technicien·ne·s, s'oppose au point de vue anthropocentré adopté par les scientifiques. De fait, dans le domaine de la recherche sur les technologies pour l'éducation, les notions d'appropriation et d'usage occupent une place prépondérante.

En parallèle à la diffusion des technologies pour l'enseignement-apprentissage, l'usage a été théorisé selon plusieurs ancrages théoriques, et la notion de non-usage a fait l'objet de nombreuses publications scientifiques [8]. Parmi la littérature consacrée aux relations que les act·eur·rice·s de l'enseignement-apprentissage entretiennent, et n'entretiennent pas, avec les objets numériques mis à leur disposition, Papi relève que pour certains représentants de la sociologie des usages « le non-usage serait davantage un choix qu'une fatalité […] un non-usage voulu […] une résistance » [9] ; pour Simonian et Audran [10], selon une approche anthropo-écologique, comme l'usage d'une technologie est motivé par son utilité, le non-usage est une absence de perception de cette utilité, qui provoque parfois le recours à d'autres outils perçus comme plus aptes à répondre aux besoins des apprenant·e·s. Selon l'anthropologie

des usages de Plantard [11], [12], les usages du numérique se construisent au cours de « processus internes », suivant trois modes : les usager·e·s bricolent alors qu'ils réalisent des tâches variées à partir des instruments dont ils disposent ; ils butinent, alors qu'ils créent de façon hasardeuse, intuitive et émotive avec les technologies ; suivant de Certeau (1980), ils braconnent, lorsque, collectivement, ils détournent « l'offre sociotechnique », en s'appuyant sur leurs « capacités de micro-résistances ».

Étudiés d'un point de vue anthropocentré, à l'opposé de la « vision mécaniste et positiviste de la technique » définie par Millerand (1999) et qui est dominante [3], les non-usages des technologies pour apprendre révèlent que le sujet choisit d'être un usager ou de ne pas l'être, selon ses croyances et selon ses besoins, et qu'il est en mesure de concevoir d'autres usages pour atteindre ses objectifs. Suivant Plantard [11], [12], l'adaptabilité, la créativité et « l'intelligence collective » sont essentielles dans les processus de construction des usages du numérique. Ceci nous amène à affirmer avec Perriault que « l'individu détient fondamentalement une part de liberté dans le choix qu'il fait d'un outil pour s'en servir conformément ou non à son mode d'emploi » [13], et à émettre l'hypothèse que la profusion des ressources induite par les technologies numériques est susceptible de participer au développement de nouvelles compétences pour les sujets, telles que l'adaptabilité, la créativité et l'intelligence collective.

4 Constructions Collectives et Singularité des Parcours

La diversification des possibilités de choix portée par l'innovation technologique, aussi féconde que potentiellement problématique, est au cœur de la conceptualisation que nous avons proposée ailleurs des « connexions démocratiques » [14], [15], [16], comme manière la plus riche heuristiquement de penser le déploiement de nombre de relations interindividuelles aujourd'hui, dans la vie privée comme dans la vie professionnelle et/ou politique. Le monde numérique semble en effet pouvoir être considéré comme la traduction technique et concrète de l'idéal théorique et abstrait du lien dans une société individualiste. Profusion, disponibilité, interactivité, voire non-usage, ouvrent sur un vaste champ de possibles, en particulier en termes de parcours de formation, sur la base du seul consentement, dans une double perspective d'individuation et d'individualisation [17]. La manière dont chacun se saisit des différentes innovations technologiques, en particulier numériques, est désormais l'un des points d'appui et de positionnement par lequel l'individu se construit lui-même et comme être social. Par corollaire, la construction professionnelle et scientifique du·de la doctorant·e passe pour une part significative par de tels choix, qui peuvent être déterminants à la fois pour la constitution d'un réseau de collaboration, que pour une ouverture à l'international, ou l'ancrage dans une communauté de recherche (épistémologique et/ou problématique). L'enjeu est donc bien, pour comprendre les mutations anthropologiques à la fois révélées et induites par le déploiement du numérique (et pour comprendre la manière dont l'expérience doctorale est reconfigurée à cette aune), de penser avant tout ce qu'implique et nous enseigne l'« appropriation sociale d'une technologie » [18].

L'utilisation de quelques outils numériques dans le cadre de la vie quotidienne d'un labo junior comme le nôtre permet déjà de percevoir un grand nombre de choix de positionnements et d'usages chez les quelque soixante-dix doctorant·e·s et jeunes docteur·e·s concernés. De simples constats quotidiens, comme l'utilisation inégale d'outils de travail en commun ou de programmation de celui-ci (Framapad, Doodle, Hypothèses, etc.), ou les rapports différents de chacun à la réception et à l'émission sur des listes de diffusion de courriels internes au groupe des doctorant·e·s, témoignent de la diversité des choix faits par chacun et de la manière dont ils peuvent s'agréger. Indépendamment de la distance physique qu'il existe entre chacun du fait de ses lieux de vie et de travail, des binômes, voire des groupes plus étendus de travail sur une publication scientifique ou l'organisation d'un événement à l'échelle du laboratoire se forment. Certains s'approfondissent et se prolongent au-delà de ce qu'il pouvait apparaitre comme possible à leur commencement, d'autres restent à l'état de proclamations initiales. Fluidité et adaptabilité en fonction des obligations pratiques et avancées respectives de chacun dans son propre travail de thèse semblent en être les maitres-mots.

5 Conclusion

Le foisonnement des ressources rendues disponibles par les technologies numériques, offre de situer la formation doctorale en sciences humaines et sociales comme un parcours qui se réalise parmi les choix. Dans un monde numérique, les doctorant·e·s acquièrent au cours de ce parcours des compétences nouvelles, telles l'adaptabilité, la créativité et l'intelligence collective. Dans l'ensemble des cas envisagés, il semble que la plus large palette de choix offerts quant à la manière, la fréquence et l'optique dans laquelle les doctorant·e·s se connectent, entre autres, les rendent davantage à même de façonner leur propre expérience doctorale dans sa singularité, sans pour autant modifier radicalement la nature même des défis qu'ils ont à relever durant cette période. En un mot, cela rend cette expérience de formation potentiellement plus riche, mais aussi plus complexe à gérer et organiser.

Références

1. Gauchet, M. *La condition historique*. Paris : Gallimard. (2003).
2. Albero, B., Simonian, S., & Eneau, J. *Des humains & des machines : Hommage aux travaux d'une exploratrice*. Dijon : Éditions Raison et Passions, (2019).
3. Fluckiger, C. Numérique en formation : Des mythes aux approches critiques. *Éducation permanente, (219)*, (2019), 19-30.
4. Gérard, L. *Le doctorat : un rite de passage. Analyse du parcours doctoral et post-doctoral*. Paris : Téraèdre. (2014).
5. Jorro, A. Développement professionnel. Dans A. Jorro (Éd.), *Dictionnaire des concepts de la professionnalisation* (p. 75-79). Louvain-la-Neuve : De Boeck Supérieur. (2013).
6. Flichy, P. Internet ou la communauté scientifique idéale. *Réseaux. Communication - Technologie - Société, 17(97)*, (1999), 77-120.

7. Bernard, F.-X., & Fluckiger, C. Innovation technologique, innovation pédagogique. Éclairage de recherches empiriques en sciences de l'éducation. *Spirale, (63)*, (2019) 3-10.
8. Trestini, M. (2012). Causes de non-usage des TICE à l'Université : Des changements ? Cas des enseignants du supérieur, *Recherches & éducations, (6)*, 15-33.
9. Papi, C. Causes et motifs du non-usage de ressources numériques. Logiques d'usage des étudiants en formation initiale. *Recherches & éducations, (6)*, 127-142. (2012).
10. Simonian, S., & Audran, J. Approche anthropo-écologique du non-usage. Le cas des outils communicationnels des plateformes d'apprentissage en ligne. *Recherches & éducations, (6)*, 161-177. (2012).
11. Plantard, P. Temps numériques et contretemps pédagogiques en Collège Connecté. *Distances et médiations des savoirs (16)*. (2016).
12. Plantard, P. Mythes, rites et tribus du numérique. *Éducation permanente, (219)*, 116-127. (2019).
13. Perriault, J. *La logique de l'usage. Essai sur les machines à communiquer*. Paris : L'Harmattan. (2008).
14. Roelens, C. Les connexions démocratiques : influences consenties et légitimité. Rendre l'individu auteur. Actes du colloque LUDOVIA 2018 « Innovation(s) / Institution(s) du numérique ». En ligne : http://culture.numerique.free.fr/publications/ludo18/Roelens_Ludovia_2018.pdf). Ax-les-Thermes. (2018).
15. Roelens, C. Individualisme contemporain et distance : appropriations des technologies numériques et médiation des savoirs. Actes du colloque international ÉDUCATION 4.1 ! Distances, médiations des savoirs et des formations, CNED Poitiers, En ligne : https://education4-1.sciencesconf.org/215435/document. (2019).
16. Roelens, C. *L'autorité bienveillante dans la modernité démocratique : entre éducation, pédagogie et politique*. Thèse de doctorat en sciences de l'éducation, UJM Saint-Etienne. (2019).
17. Gauchet, M. *Le nouveau monde. L'avènement de la démocratie IV*. Paris : Gallimard. (2017).
18. Gauchet, M. Conclusion : vers une mutation anthropologique ? (Entretien avec Nicole Aubert et Claudine Haroche). Dans N. Aubert, *L'individu hypermoderne* (pp. 405-420). Toulouse : Erès. (2004/2017).

2.6 La confiance et son impact dans le domaine du numérique

José Samaniego, Stéphanie Mailles-Viard Metz, Nathalie Blanc

La confiance et son impact dans le domaine du numérique

José Samaniego[12], Stéphanie Mailles-Viard Metz[1], Nathalie Blanc[2]

[1] Université Paul Valéry, CNRS Praxiling UMR 5267, Route de Mende, Montpellier, France
j.samaniegocho@gmail.com
[2] Université Paul Valéry, Epsylon, EA 4556, Rue du Professeur Henri Serre, Montpellier, France

Résumé. Dans cette étude, nous nous sommes intéressés à évaluer la confiance généralisée, qui caractérise la propension d'un individu à attribuer sa confiance avec l'objectif d'examiner si cette confiance généralisée peut être amenée à varier selon le contexte dans une situation d'interaction médiatisée par l'ordinateur. Pour ce faire, nous avons repris l'expérience de Paxton et Glanville (2015) et proposé aux participants le « Jeu de la Confiance » (Berg et al., 1995) en faisant artificiellement varier l'environnement (Confiant vs Non confiant). Le score de confiance généralisée des participants au début de la phase test nous a permis de les catégoriser. Nos résultats montrent qu'il n'y a pas de différence entre les niveaux de confiance généralisée au cours de différents temps de mesure, quel que soit le groupe. Ils pourraient permettre de concevoir des outils d'enseignement adaptés susceptibles de prendre en compte la confiance envers l'enseignant ou l'outil dans le cadre du travail collaboratif.

Mots-clés. Confiance, émotion, numérique, environnement, apprentissage

Abstract. In this study, we attempted to assess the generalized trust, which characterizes an individual's propensity to attribute trust with the objective of examining whether this generalized trust may vary by the context in a situation of interaction mediated by the computer. To do this, we took the previous research of Paxton and Glanville (2015) and presented the participants the « Trust Game » (Berg et al., 1995), artificially varying the environment (Trust vs. Non-Trust). The generalized trust score of the participants at the beginning of the test phase allowed us to categorize them. Our results show that there is no difference between the levels of generalized trust[i] over different measurement times, regardless of the group. They could be used to design adapted teaching tools that can take into account the trust towards the teacher or the tool in the context of collaborative work.

Keywords. Trust, emotion, digital, environment, learning

1 Introduction

La notion de confiance, du latin confidere : cum, « avec » et fidere « fier », peut paraître simple à appréhender, intuitive, et présente dans le quotidien de tous. De nombreuses disciplines telles que la sociologie, le marketing, la philosophie, la finance ou encore la psychologie (pour n'en citer que quelques-unes) portent un regard spécifique sur la notion de confiance, ce qui rend complexe la possibilité de trouver un consensus. Parmi les nombreux auteurs [1] qui ont tenté de trouver un point de convergence parmi les dizaines de définitions possibles de ce terme (plus de 70) nous retenons en particulier celle de Mayer, Davis et Schoorman [2] qui l'envisagent comme la « volonté d'une personne d'être vulnérable aux actions d'une autre personne en ayant la conviction que cette dernière effectuera une action importante pour le croyeur, indépendamment de l'habilité à contrôler l'autre tiers ». En effet, cette définition, qui se veut générale et consensuelle, implique qu'on remet quelque chose de précieux à quelqu'un, en se fiant à lui et en s'abandonnant ainsi à sa bienveillance et à sa bonne foi. Aujourd'hui, pour simplifier nos liens de communication et donc nos relations, nous sommes exposés en permanence à internet, ce qui nous amène à ce que les spécialistes désignent aujourd'hui comme la 4ème révolution industrielle (Internet of Things ou IoT). Cette dynamique contemporaine n'est pas sans conséquence sur le fonctionnement psychologique, et ces nouveaux usages nous amènent légitimement à nous interroger : peut-on encore parler de confiance ?

2 Cadre théorique

Ben Naoui et Zaiem [3] ont tenté de trouver un lien entre la confiance dite « classique » et la confiance « en ligne / e-confiance », soit cette même confiance médiée par une interface numérique connectée à internet. A travers une méta-analyse, ils ont pu mettre en évidence des régularités : toutes les approches considèrent la confiance lors de situations incertaines ou risquées, il y a toujours une personne qui accorde sa confiance, appelé « trustor », et un objet de la confiance, appelé «trustee ».

Le « trustee » peut être un individu, un groupe d'individus, une organisation ou une institution (exemple de l'internet, de la technologie, etc.). La plupart des recherches relatives au concept de l'e-confiance [4], [5], [6], [7], [8] se focalise sur les transactions en ligne.

Néanmoins, l'outil Internet ne se résume pas uniquement au commerce électronique. Bien au contraire, Internet est un média incontournable de communication qui peut contenir aussi bien des sites à vocation commerciale ou informative. Il s'avère donc essentiel de tenir compte de la nature du site (réseau social, module e-learning…) dans l'étude de la confiance des utilisateurs. De plus en plus, les formations académiques prennent un tournant et s'orientent vers la possibilité d'un format à distance, avec la volonté de s'adapter au profil des apprenants. Toutefois, nous n'avons pas connaissance de formations à distance qui prendraient en considération le niveau de confiance des utilisateurs. Weckert [7] transpose la confiance dite classique à la confiance dans le domaine du numérique dans toutes ses dimensions. Parmi les différents niveaux de confiance, on s'intéresse en particulier à la propension à la

confiance [2], [9] ou confiance généralisée [10] assimilable à un trait de personnalité (ou confiance-trait). Celle-ci consiste en une prédisposition généralisée à faire confiance, quel que soit le contexte rencontré ou les individus concernés. Elle est durable dans le temps et basée sur des expériences de vie mais aussi sur le tempérament de l'individu, ainsi que sur des facteurs génétiques et biologiques dans une certaine mesure. Selon Mayer et al. [2], cette « volonté générale de faire confiance aux autres » augmente le niveau mesurable de confiance avant même d'avoir des informations sur le trustee. Elle est typiquement mesurée à travers l'échelle RTS (Rosenberg Trust Scale) proposée par Rosenberg [11], qui constitue le « golden standard » de cette mesure. Par ailleurs Rosenberg a défini un seuil de catégorisation pour cette échelle : en dessous de 6, l'individu possède un niveau de CG bas et au-delà de 6 un niveau de CG haut. Frazier [12] a récemment proposé une nouvelle version de cette échelle appelée TPS (Trust Propensity Scale), qui est plus pertinente et robuste encore que la précédente.

Paxton et Glanville [13] ont récemment introduit la notion d'une confiance généralisée (CG) dite malléable, qui s'adapterait en fonction des contraintes de l'environnement, valable également pour l'e-confiance. Cette notion s'avère essentielle dans notre vie quotidienne car elle pourrait indiquer que chaque expérience vécue pourrait avoir un impact conséquent sur notre vision de la confiance. Une expérience qui mettrait à l'épreuve notre confiance pourrait impacter durablement notre confiance généralisée. En effet, en fonction de la nature des interactions, les auteurs observent une altération du niveau de CG. On parle dans ce cas de confiance-état, ce qui contredit la notion de confiance-trait [9]. Les auteurs rapportent des résultats d'autant plus importants lorsque l'environnement de confiance n'est pas congruent ou n'est pas en accord avec le niveau de CG de l'individu.

Dans cette étude, nous avons souhaité mettre à l'épreuve ces deux notions en répliquant l'étude de Paxton et Glanville [13], sachant que nous avons utilisé comme ces auteurs le jeu de la confiance* [14] mais dans une version modifiée** que nous avons développé au regard de leurs recommandations, à savoir : préciser un nom pour les autres joueurs, ajouter une photographie et prélever des mesures de la confiance généralisée à différents temps au cours de l'expérimentation .Nous avons également ajouté l'utilisation de l'échelle de Frazier [12] ainsi que des jetons pour fournir un support physique à la gestion des sommes d'argent.

Notre plan expérimental était composé de trois variables indépendantes, la première étant le « niveau de CG » correspondant à deux modalités : CG Basse ou CG- (score de 5 ou moins dans l'échelle RTS) et CG Haute ou CG+ (score au-dessus de 5). La deuxième variable indépendante était le facteur « Environnement » correspondant à deux modalités : Environnement Confiant (E+) et Environnement Non Confiant (E-). La troisième variable indépendante était le facteur « Temps », à cinq modalités : T0, T1, T2, T3, T4 et T5 et correspondait aux différents temps de mesure de la CG, soit en début de passation (T0), en fin de passation (T1), 3 jours plus tard (T2), 7 jours plus tard (T3), 14 jours plus tard (T4) et 31 jours plus tard (T5). Nous avons choisi de faire varier ces facteurs de manière à avoir quatre conditions : Notre variable dépendante était le score aux échelles de CG. Nous avons donc les quatre groupes suivants : « CG+ E+ », « CG+ E-», « CG- E-» et « CG- E+ » avec les deux mesures des échelles différentes.

*Le jeu de la confiance consiste à prendre des décisions économiques : donner de l'argent, en renvoyer ou le garder. Les joueurs sont associés par paires anonymes. Le participant possède une somme de base et peut être affecté à la place de premier ou deuxième joueur. 1er joueur: décide d'envoyer une somme d'argent à un autre participant, en sachant que la somme qu'il envoie sera systématiquement triplée par l'expérimentateur. 2e joueur: reçoit la somme du premier joueur et décide de renvoyer de l'argent ou de garder la somme.

**Les auteures induisent un environnement en présentant un tableau où les autres paires de joueurs ont donné plus ou moins d'argent, ainsi que la réponse associée. Le comportement du joueur 1 est aussi adapté en fonction de la condition. Dans un environnement confiant, le premier joueur donne une somme importante et le deuxième en revoie encore plus. Dans le cas contraire, le premier envoie peu ou pas d'argent et le deuxième n'en revoie quasiment jamais.

Nous nous attendons à retrouver les résultats de Paxton et Glanville, soit à avoir une CG malléable, mais également à une durabilité dans le temps des variations du niveau de CG. Dans un environnement confiant, on s'attend à observer une hausse du niveau de CG et dans un environnement méfiant, on s'attend à observer une baisse de ce niveau. Les effets devraient être encore plus marqués en cas de non-congruence de l'environnement avec le niveau de CG initial.

3 Méthode

3.1 Participants

Soixante-quatorze participants (48 hommes / 26 femmes, âge moyen: 20 ans) ont volontairement pris part à cette étude. Tous ont rempli un formulaire de consentement, sachant que leur recrutement s'est déroulé au sein de l'IUT de Montpellier-Sète. Selon leur niveau de CG, les participants étaient affectés à l'un des deux environnements (confiant vs méfiant), cette affectation ayant fait l'objet d'un contrebalancement pour favoriser une répartition équilibrée des différents profils de CG.

3.2 Matériel

Nous avons utilisé les échelles TPS et RTS pour mesurer le niveau de confiance généralisée et le jeu de la confiance dans sa version informatisée et modifiée. Les participants disposaient du prénom de l'autre joueur, d'une photo et de la somme envoyée. Nous avons utilisé des jetons de poker pour fournir un support physique à l'argent virtuel dont disposaient les participants. Pour les questionnaires de suivi et de mesure du niveau de CG, nous avons utilisé les logiciels CloudHQ pour les envois programmés et Google Forms où la question contrôle « Avez-vous récemment vécu un évènement qui a pu modifier votre perception des autres ou brisé votre confiance ? » était intégrée.

3.3 Procédure

Nous avons utilisé la plateforme biométrique iMotions®[15] dans sa version 7.2, installée dans un ordinateur portable HP ayant les caractéristiques techniques recommandées pour présenter nos stimuli et collecter des données. Nous avons utilisé la fonction "Survey" de ce logiciel pour conduire cette expérimentation, en présentant des échelles de Likert allant de 0 à 10 pour les mesures de CG, ainsi que pour les sommes d'argent variables qui nécessitaient de la prise de décision. Chaque passation avait une durée de 15 minutes environ. Après avoir signé le formulaire de consentement, le niveau de CG des participants était évalué. Les instructions et les règles du jeu leur étaient ensuite précisées via un écran d'introduction à l'expérimentation. Nous leur avons demandé d'être le plus naturel possible, comme s'ils étaient en situation de face à face avec l'autre joueur tout en rappelant que nous étions à côté en cas de problème ou de question. Les participants disposaient d'un tableau récapitulatif des 10 tours qui avaient été joués avant qu'ils ne rejoignent la partie, puis commençaient le jeu.

Les participants étaient placés systématiquement à la place du joueur 2 (ils devaient donc prendre la décision de renvoyer de l'argent ou le garder). Selon l'environnement dans lequel ils étaient placés, ils pouvaient recevoir (environnement confiant) ou pas (environnement non-confiant) de l'argent du premier joueur. S'ils ne recevaient pas d'argent du premier joueur, le jeu ne leur permettait pas d'en envoyer. A la fin de chaque tour, un tableau récapitulatif résumait les actions des autres paires de joueurs pendant le dernier tour. Cette procédure se répétait ainsi pendant 10 tours au total, jusqu'à atteindre un dernier tableau récapitulatif des dix tours joués présenté en fin de passation. Ensuite, nous avons envoyé l'échelle de CG aux participants dans les différents temps de mesure (T2, T3, T4 et T5) systématiquement accompagnée de la question contrôle.

4 Résultats

Pour réaliser nos analyses statistiques, nous avons utilisé le logiciel RStudio [16] version 3.5.3. Nous avons tout d'abord étudié les effets simples au sein de chaque condition, en comparant les moyennes des scores aux échelles de CG à T0 et à T1 (voir Figures 1 et 2).

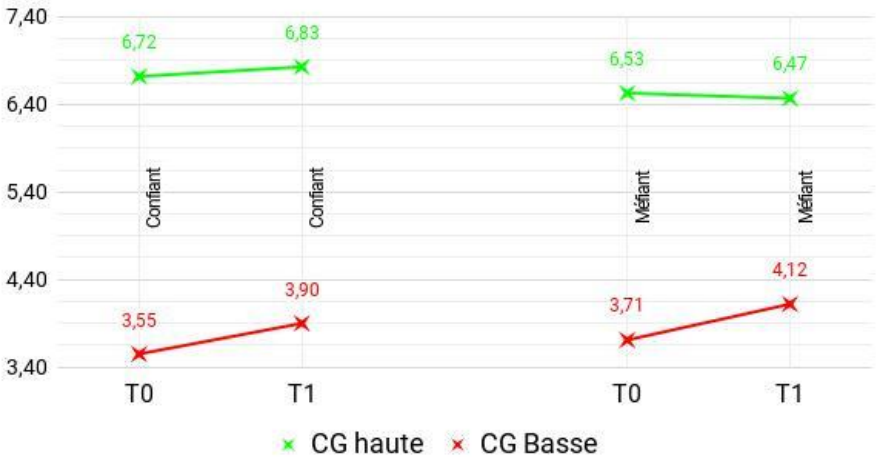


Fig. 1. Score à l'échelle RTS en fonction de la condition.

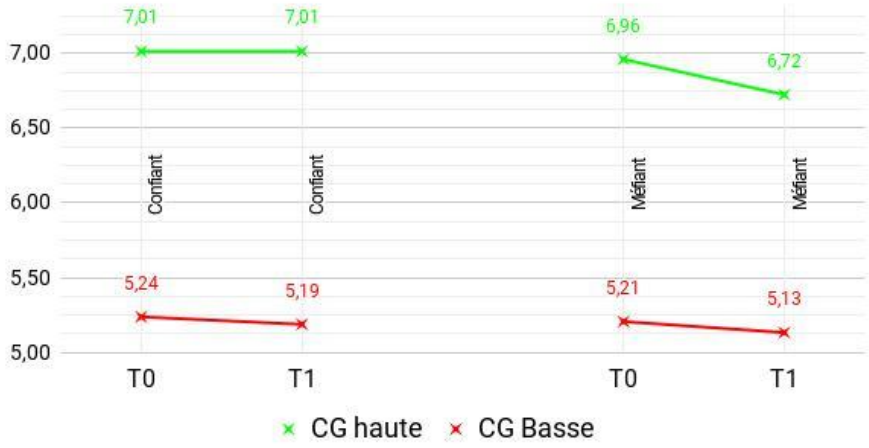


Fig. 2. Score à l'échelle TPS en fonction de la condition.

Pour les moyennes des échantillons suivant une loi normale, nous avons utilisé le test t de Welch et pour les moyennes ne suivant pas une loi normale nous avons utilisé un test de rangs signés de Wilcoxon. Pour le groupe « CG+ E+ », nous n'avons pas trouvé de différence entre T0 et T1 pour l'échelle RTS : $Z = 142.5$, $p > 0.5$ ni pour l'échelle TPS : $t(33.8) = 0$, $p > 1$. Pour le groupe « CG- E+ », nous n'avons pas trouvé de différence entre T0 et T1 pour l'échelle RTS : $Z = 188.5$, $p > 0.7$ ni pour l'échelle TPS : $t(37.7) = 0.07$, $p > 0.9$. Pour le groupe « CG- E- », nous n'avons pas trouvé de différence entre T0 et T1 pour l'échelle RTS : $Z = 129$, $p > 0.5$ ni pour l'échelle TPS : $t(31.9) = 0.1$, $p > 0.9$. Pour le groupe « CG+ E- », nous n'avons pas trouvé de différence entre T0 et T1 pour l'échelle RTS : $Z = 185.5$, $p > 0.88$ ni pour l'échelle TPS : $t(36) = 0.49$, $p > 0.6$.

Nous avons également effectué une analyse de la variance (ANOVA) avec comme variable aléatoire le « niveau de CG », comme première variable indépendante le facteur « Environnement » en inter-sujet et comme deuxième variable indépendante le facteur « Temps » (i.e T0, T1, T2, T3, T4 et T5) en intra-sujet. La variable dépendante étudiée était le score aux deux échelles de CG (RTS et TPS). Nous avons trouvé un effet principal de la variable « niveau de CG » sur le score à l'échelle de CG TPS : $F(3,387) = 46.5$, $p = 2.2^{e}\text{-}16$ (voir Figure 3)

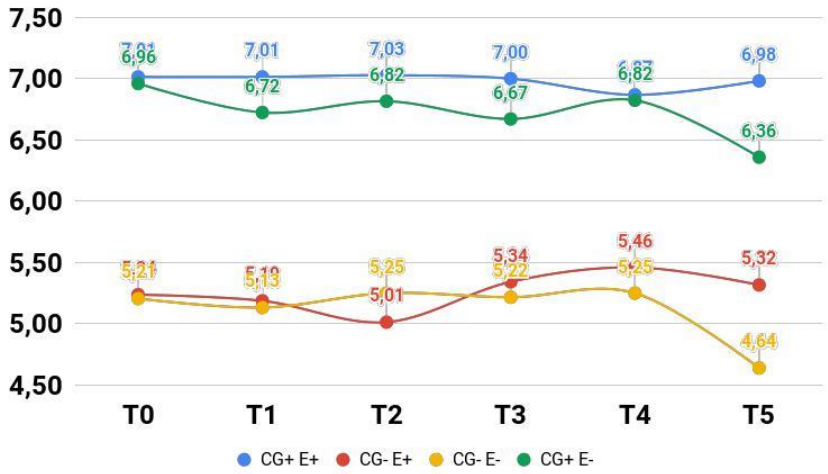


Fig. 3. Evolution des scores de CG en fonction du temps, échelle TPS.

et sur l'échelle RTS : $F(3,387) = 150$, $p < 2.2^{e}$-16 (voir Figure 4).

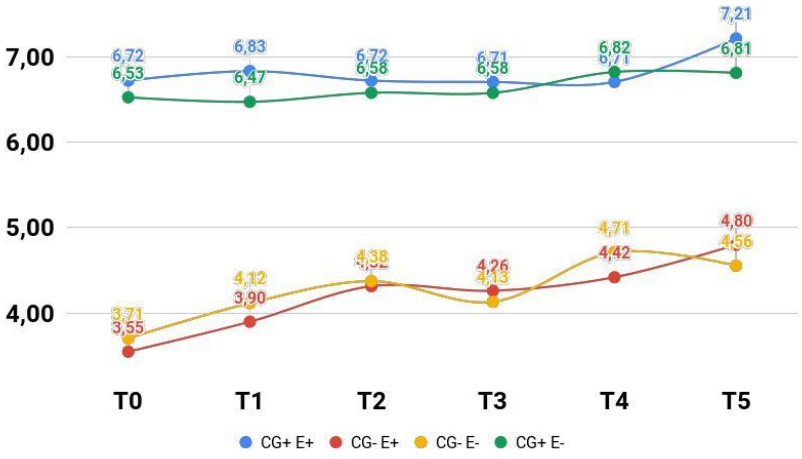


Fig. 4. Evolution des scores de CG en fonction du temps, échelle RTS.

La variable « Temps » est un fort prédicteur sur les scores à l'échelle RTS : $F(5,387) = 3.92$, $p < .002$. Nous avons trouvé que le score à l'échelle TPS était un fort prédicteur du score à l'échelle RTS : $F(1,387) = 295.7$, $p < 2^e$-16 et que le score à l'échelle RTS était un fort prédicteur du score à l'échelle TPS : $F(1,387) = 298.1$, $p < 2.2^e$-16. Nous n'avons pas trouvé d'effet d'interaction.

Nous avons effectué une deuxième ANOVA identique à la précédente en retirant cette fois-ci le score aux échelles de CG à T5. Nous avons trouvé les mêmes résultats que lors de la première ANOVA sauf pour la variable « Temps » qui n'a plus d'influence significative sur les scores à l'échelle RTS : $F(4,336) = 2.21$, $p > .05$.

5 Discussion

L'objectif de notre étude était de répliquer les résultats rapportés par Paxton et Glanville [13] tout en tenant compte de leurs recommandations. Nos prédictions étaient que les participants avec un niveau de CG préalable non congruent avec l'environnement auquel ils sont exposés (CG Basse + environnement confiant ou CG Haute + environnement méfiant) devraient présenter un changement significatif de leur score aux échelles de CG, d'autant plus que ce changement serait durable dans le temps (jusqu'à 31 jours plus tard). Nos résultats ont montré qu'en fonction du niveau de CG initial, quelle que soit la condition d'affectation et son caractère congruent ou non-congruent, et quel que soit le temps de mesure, le score aux échelles de CG TPS reste identique. Pour l'échelle RTS en revanche, le score est variable en fonction du temps de mesure. Plusieurs explications sont possibles pour ce résultat : il pourrait aller dans le sens de Paxton et Glanville et indiquer la malléabilité de la CG, être dû à une qualité de mesure moins importante [12] de l'échelle RTS en comparaison à l'échelle TPS ou tout simplement signifier que ces échelles ne mesurent pas la confiance généralisée de la même manière. Bien qu'elles mesurent le même concept, la TPS pourrait être davantage orientée vers une mesure stable, semblable à celle d'un trait, tandis que la RTS pourrait être davantage une mesure semblable à un état. Toutefois lorsque nous avons retiré la mesure au T5 (étant donné que nous avons perdu 27% de nos participants à ce temps de mesure), l'effet du temps sur le score de CG n'était plus significatif. La variabilité était probablement expliquée par un manque de représentativité de l'échantillon restant à T5. Suite à cette correction et à la robustesse des effets, une explication possible de ces résultats pourrait être que la confiance généralisée n'est pas malléable, comme l'ont conclu Paxton et Glanville [13], mais pourrait être plus stable, à l'image d'un trait de personnalité [9], [17]

Un autre apport de cette étude est que nous observons une forme de continuité dans le niveau de confiance généralisée, et que ce niveau reste cohérent avec la distinction initiale qui a permis de créer les différents groupes (seuil d'attribution). Par exemple, les groupes « CG+ E+ » et « CG+ E- », ayant une CG élevée initialement, ne descendent jamais en dessous du seuil de 6 et inversement pour les groupes « CG- E » et « CG- E+ » dont la CG était basse initialement et qui n'atteignent jamais le score de 6. Cette distinction s'est avérée pertinente et va dans le sens des propos de Rosenberg [11].

6 Conclusion

La confiance est considérée comme un processus d'apprentissage [18], apprise à partir d'interactions et discussions [19]. Au même titre que les autres types d'apprentissage, la confiance se produit lorsque les conditions sont favorables et permet d'accentuer les phénomènes de coopération [20], [21]. Elle est considérée comme la pierre angulaire des bonnes relations à long terme [22], serait à l'origine d'une augmentation de la qualité de vie, au niveau biologique et social et de la satisfaction professionnelle, l'engagement, la productivité et le partage de connaissances [23], [24]. La diminution des facteurs liés au stress (baisse de la méfiance), du coût énergétique (interactions plus fluides et qui ne nécessitent pas d'une analyse extrêmement poussée) permettrait une meilleure utilisation du temps et l'établissement de liens [25], [24] plus robustes dans le temps. Dans une perspective pédagogique, nos résultats pourraient s'appliquer à la conception d'outils d'enseignement qui seraient susceptibles de prendre en compte la confiance envers l'enseignant, l'outil ou envers les autres utilisateurs dans le cadre du travail collaboratif. La confiance est un élément central dans le tryptique apprenant-outil-enseignant qui n'est cependant pas encore abordé mais dont les futures recherches en pédagogie pourraient s'intéresser. Dans une perspective future, nous souhaiterions aborder la question de l'importance des émotions, relevée par Gillespie et Mann [26], qui ont proposé l'existence d'une confiance affective. De manière générale, les émotions influenceraient la confiance [27], en particulier les émotions telles que la joie et la satisfaction [28]. La mesure des paramètres physiologiques (GSR, BPM, Température et diamètre pupillaire) peut apporter des informations concernant les changements d'état émotionnel et subjectif des individus lors de tâches cognitives [29], [30].

Remerciements. Nous souhaiterions remercier l'IUT de Montpellier-Sète pour nous avoir accordé l'usage de leurs locaux pour cette expérience, le laboratoire Epsylon pour l'utilisation des box expérimentaux, les nombreux étudiants qui se sont portés volontaires ainsi qu'à la société iMotions pour le prêt du logiciel. Nous remercions également les partenaires du projet E-Confiance: l'Université Paul Valéry Montpellier 3, l'Université de Montpellier, le CNRS, la société Pikcio, la Région Occitanie, l'Union Européenne, le laboratoire Epsylon et le laboratoire Praxiling. Ce travail de

recherche a été financé par le FEDER (Fonds Européen de Développement Régional) et la Région Occitanie, contrat n°166795.

Références bibliographiques

1. Bauer, P. C. (2014). Conceptualizing and measuring trust and trustworthiness. Political Concepts: Committee on Concepts and Methods Working Paper Series, 61, 1-27.

2. Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. Academy of management review, 20(3), 709-734.

3. Naoui, F. B., & Zaiem, I. (2015). Reasons for consumer trust in health websites: An approach integrating website-based factors and personality-based factors. International Journal of Management Sciences and Business Research.

4. Doney, P. M., Cannon, J. P., & Mullen, M. R. (1998). Understanding the influence of national culture on the development of trust. Academy of management review, 23(3), 601-620.

5. Jarvenpaa, S. L., Tractinsky, N., & Saarinen, L. (1999). Consumer trust in an Internet store: A cross-cultural validation. Journal of Computer-Mediated Communication, 5(2), JCMC526.

6. McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). The impact of initial consumer trust on intentions to transact with a web site: a trust building model. The journal of strategic information systems, 11(3-4), 297-323.

7. Weckert, J. (2005). Trust in cyberspace. The impact of the internet on our moral lives, 95-117.

8. De Vries, P. (2006, May). Social presence as a conduit to the social dimensions of online trust. In International Conference on Persuasive Technology (pp. 55-59). Springer, Berlin, Heidelberg.

9. Uslaner, E. M. (2002). The Moral Foundations of Trust. Cambridge, UK: Cambridge University Press.

10. Yamagishi, T., and Yamagishi, M. (1994). Trust and Commitment in the United States and Japan.' Motivation and Emotion 18(2):129–66.

11. Rosenberg, M. (1956). Misanthropy and political ideology. American Sociological Review, 21, 690–695.

12. Frazier, M. L., Johnson, P. D., & Fainshmidt, S. (2013). Development and validation of a propensity to trust scale. Journal of Trust Research, 3(2), 76-97.

13. Paxton, P., & Glanville, J. L. (2015). Is trust rigid or malleable? A laboratory experiment. Social Psychology Quarterly, 78(2), 194-204.

14. Berg, J., Dickhaut, J., & McCabe, K. (1995). Trust, reciprocity, and social history. Games and economic behavior, 10(1), 122-142.

15. iMotions Global. (n.d.). Facial Expression Analysis Solutions. Retrieved September 14, 2019, from https://imotions.com/facial-expressions/

16. RStudio Team (2018). RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL http://www.rstudio.com/.

17. Uslaner, E. M. (2008). The Foundations of Trust: Macro and Micro. Cambridge Journal of Economics 32:289–94.

18. Sabel, C. F. (1993). Studied trust: Building new forms of cooperation in a volatile economy. Human relations, 46(9), 1133-1170.

19. Powell, W. W. (1996). Trust-Based Forms of Govornanco. Trust in organizations: Frontiers of theory and research, 51.

20. Dyer, J. H., & Singh, H. (1998). The relational view: Cooperative strategy and sources of interorganizational competitive advantage. Academy of management review, 23(4), 660-679.

21. Xie, X., Yuan, T., Zhou, X., & Cheng, X. (2018). Research on trust model in container-based cloud service. Computers, Materials and Continua, 56(2), 273-283.

22. Wu, G., Hu, X., & Wu, Y. (2010). Effects of perceived interactivity, perceived web assurance and disposition to trust on initial online trust. Journal of Computer-Mediated Communication, 16(1), 1-26.

23. Kramer, R. M. (1999). Trust and distrust in organizations: Emerging perspectives, enduring questions. Annual review of psychology, 50(1), 569-598.

24. Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. Management science, 50(11), 1477-1490.

25. Hansen, M. H., Morrow Jr, J. L., & Batista, J. C. (2002). The impact of trust on cooperative membership retention, performance, and satisfaction: an exploratory study. The International Food and Agribusiness Management Review, 5(1), 41-59.

26. Gillepsie, N., & Mann, L. (2004). Transformational leadership and shared values: The building blocks of trust. Journal of Managerial Psychology, 588-607.

27. Dunn, J. R., & Schweitzer, M. E. (2005). Feeling and believing: the influence of emotion on trust. Journal of personality and social psychology, 88(5), 736.

28. Bewsell, G. R. (2012). Distrust, fear and emotional learning: An online auction perspective. Journal of theoretical and applied electronic commerce research, 7(2), 1-12.

29. Crosby, L. A., & Johnson, S. L. (2001). Technology: Friend or foe to customer relationships?. Marketing Management, 10(4), 10.

30. Della Porta, D. (2012). Critical trust: Social movements and democracy in times of crisis. Cambio. Rivista sulle Trasformazioni Sociali, 2(4), 33-43.

3

Axe 3 : Impacts sociétaux des nouveaux processus d'acquisition : construction, co-construction et modélisation des savoirs.

3.1 Quelle maîtrise dans les milieux de savoir du digital ?

Olivier LE DEUFF, MICA, Université Bordeaux Montaigne

Cette conférence a pour but d'interroger les conditions d'une maîtrise dans le contexte des milieux de savoir des environnements digitaux. L'idée de maîtrise à évolué d'une vision procédurale et contrôlée qui consistait notamment à maîtriser l'information. Il est désormais impossible de tout maîtriser, et c'est justement cette prise de conscience qui permet la définition d'une nouvelle forme de maîtrise. Nous voulons montrer que cette maîtrise suppose l'inscription dans des traditions qui sont celles du travail intellectuel et qu'il s'agit de discerner les continuités plutôt que des ruptures pour comprendre les enjeux actuels en matière de formation et de constitution des savoirs alors que des changements de régime documentaire semblent se produire.

Quelle maîtrise dans les milieux de savoir du digital ?

Olivier Le Deuff

Colloque Jeunes Chercheurs PRAXILING 2019 :

Représentations et transmission des connaissances à la lumière de l'innovation numérique

Jeudi 7 novembre.

Un nouveau contexte informationnel

- Entre raison graphique et raison computationnelle
- Entre web originel, web 2.0 et une mobilité accrue
- Une accessibilité technique accrue mais une accessibilité intellectuelle complexe
- Un nouveau régime documentaire
- Le besoin d'évaluer l'information

2

Milieux de savoir

- Nous préférons désormais appeler « milieux de savoir » les lieux de savoir comme les bibliothèques qui sont impactés par les humanités digitales. En effet, le concept de milieu permet d'intégrer plus aisément les aspects de **médiation**, notamment les **médiations documentaires dites numériques**, mais également le **rôle des médias** qui renouvèlent les modes d'interactions entre professionnels et usages. Ce concept de milieu prolonge donc celui de lieu de savoir (Jacob, 2007) qui met surtout en avant un espace visible, tandis que celle de milieu renvoie également au concept de **milieu associé** défini par Simondon (Simondon, 1989) et popularisé par Bernard Stiegler (Stiegler, 2008) et qui a le mérite de mieux articuler objets techniques et individus sociaux et techniques. Le milieu associé permet une évolution positive de la structure et de ses différents individus, tandis que son opposé **le milieu dissocié** opère des courts-circuits qui ne servent qu'aux desseins de **ceux qui contrôlent le milieu** au détriment de ceux qui l'utilisent. Dans l'idéal, les milieux de savoir doivent opérer comme des milieux associés qui permettent **une relative stabilité** de l'ensemble tout en **permettant son évolution et les innovations** en son sein, ce que Gilbert Simondon qualifie de **métastabilité**.

- Ref:http://bbf.enssib.fr/revue-enssib/consulter/revue-2016-03-005

3

Une maîtrise entre art et déprise

4

La fin d'un paradigme

- La fin de la « maîtrise de l'information »
- L'émergence des cultures de l'information et du numérique.
- Le développement d'une époque de « guérillas informationnelles ».

5

L'impossible maîtrise

- Première leçon : « il est impossible de tout maîtriser »
- Seconde leçon: « la maîtrise devient une forme d'art libéral qui est autant une maîtrise de savoir qu'une maîtrise de soi »
- Troisième leçon : « la maîtrise n'est pas innée, elle requiert un apprentissage »
- Quatrième leçon: « la maîtrise implique une capacité en tant qu'arme de défense intellectuelle et citoyenne. »

6

Des savoirs anciens

7

Traditions intellectuelles

- Renouer avec la skholé
- Une tradition technique
- Une tradition spirituelle
- Une tradition documentaire et analytique
- Une tradition mnémotechnique

8

Le retour des maîtres ?

« On peut observer qu'il est rare qu'un maître, dans votre jeune âge, vous apprenne à travailler. Le maître vous donne des énoncés de devoir; il apprécie, il cote vos travaux; parfois, et plus rarement à mesure que l'on monte en savoir, il vous propose des corrigés, il vous montre dans un modèle qu'il a composé lui-même ce qu'il aurait fallu faire. Mais il ne vous dit guère comment il fait; il laisse votre apprentissage au hasard ou à l'inspiration. Cette inexpérience du comment faire entre pour une vaste part dans l'impression de découragement que plusieurs retirent de leurs études. »[1]

1 Jean Guitton. *Le travail intellectuel, conseils à ceux qui étudient et à ceux qui écrivent*. Aubier, Editions Montaigne., 1951, p.7

9

Quelles maîtrises ?

10

Une maîtrise technique

- L'importance d'une culture technique (Simondon)
- Une maîtrise des sources d'information
- Une maîtrise de l'architecture de l'information

11

Une maîtrise des données et métadonnées

- Le rôle central des métadonnées
- Un déplacement historique de la production
- De nouveaux enjeux de prise de contrôle

12

Une maîtrise du web et des objets connectés

- De la recherche d'informations aux dispositifs de veille
- De la fiche à la « smart city » (Otlet)
- Qui sont désormais les antilopes dans un zoo ou les gorilles dans une cage ? (Briet, Pagès)

13

Une maîtrise des registres

- Les enjeux de « registration » (Ferraris)
- Se défendre face aux « ARMI » (Ferraris)
- Les enjeux de l'indexation

14

Des maîtres ?

15

Le maître face au marché de la frustration

- L'impossible zen ?
- La guerre des égos
- L'indignation comme principe de popularité
- Du tag au like, du like au want, du want au « majeur levé »

16

Le maître et le stultus

« Le stultus, c'est celui qui n'a pas souci de lui-même, qui est ouvert à tous les vents, au monde extérieur, qui laisse entrer dans son esprit toutes les représentations qui lui sont offertes par le monde extérieur, sans les examiner, sans savoir analyser ce qu'elles représentent. Il laisse ces représentations se mêler à l'intérieur de son propre esprit avec ses passions, ses désirs, ses habitudes de pensée, ses illusions, etc. Du coup, c'est celui qui est dispersé dans le temps, qui ne se souvient de rien, qui laisse sa vie s'écouler, qui n'essaie pas de la ramener à une unité en remémorisant ce qui mérite de l'être et qui ne dirige pas son attention, son vouloir, vers un but précis et bien fixé. Le stultus laisse la vie s'écouler sans mémoire ni volonté, change d'avis sans arrêt au lieu de penser à la vieillesse, à la temporalité de sa vie telle qu'elle doit être polarisée dans l'achèvement de soi. » [1]

1 Michel Foucault. *L'herméneutique du sujet. Cours au collège de France.* 1981-1982. Gallimard, Le Seuil, 2001, p. 127

17

Le maître de l'art de la disputatio

- Une culture de l'écoute
- L'échange d'arguments
- Des principes rhétoriques
- Une défense de la science… et de la
- …« sapience »

18

Le maître de la présence et de l'absence

- Être ou ne pas être (sur google ou chez les GAFAM)
- Le fantasme du non-indexé
- Une présence digitale (locataire ou propriétaire ? Cf. Merzeau)

19

Le maître des clefs

- Le retour de formes d'érudition savante
- Le besoin de synthèse dans des masses éparses
- Le besoin de guider et de tracer des chemins
- Donner des clefs d'accès

20

Des savoirs ou des armes

21

De la plume au clavier

- L'art d'écrire est déjà un art de programmer comme celui du discours
- Des formes écrites et pré-écrites dans nos logiciels (architextes)
- Des claviers (de la position assise à la position mobile)

22

Des savoirs techniques

- Une connaissance des codes de base
- Bâtir ses espaces informationnels
- Les enjeux du cloud
- La capacité à innover

23

Formation, information, déformation: que faut-il vraiment savoir ?

- Résister au surplus informationnel
- Accepter de découvrir encore et encore
- Expérimenter par soi-même
- Résister aux ARMI (appareils de registration et de mobilisation d'intentionnalité- Ferraris)

24

Le maître de l'équilibre face à l'infocalypse

- Pratiquer l'autocritique
- Se relire avant de publier
- Se relier longtemps après
- Résister aux enjeux de popularité
- Etudier et « bricoler » de nouveaux dispositifs
- Ne rien dire parfois...

25

Instituer ou resituer

26

Resituer avant de destituer

- Montée en puissance des industries de service en tant que nouvelles institutions;
- Ré-institutions des institutions et des acteurs du savoir
- Constituer de l'hyperlocal au global

27

Conclusion

28

Je m'étais fait un devoir scientifique d'incrédulité, mais à présent il fallait que je me méfie même des maîtres qui m'avaient appris à devenir incrédule.

Umberto ECO, *Le Pendule de Foucault*, Grasset, 1990

29

- **Coder, encoder, décoder…**
- *Il faut d'abord jouer avant de prétendre devenir maître du jeu*

30

3.2 Etude sur l'Argumentation de Collégiens au Sujet d'une Controverse Socio-scientifique Présentée en Vidéo

Julien Magnier

Etude sur l'Argumentation de Collégiens au Sujet d'une Controverse Socio-scientifique Présentée en Vidéo

Julien Magnier[1]

[1] Université Toulouse Jean Jaurès, CNRS UMR 5263, 5 allées Antonio Machiado, Toulouse, France
julien.magnier@univ-tlse2.fr

Résumé. La révolution numérique engendre une exposition massive des jeunes à une variété de contenus en ligne, notamment vidéo. Cette étude a pour objectif d'évaluer l'influence d'une exposition à une controverse socio-scientifique présentée par un ensemble de vidéos sur des productions écrites argumentées de collégiens. Nous analysons l'impact de cette exposition ainsi que d'activités en lien avec les vidéos sur la qualité des argumentations produites. Nous montrons que les collégiens formulent des réponses plus riches et prennent plus en compte les arguments opposés à leur position après avoir consulté les vidéos, et qu'une activité de tri des vidéos leur fait produire des réponses plus argumentées.

Mots-clés. Apprentissage par vidéo, Controverse socio-scientifique, Argumentation, Contre-argumentation,

Abstract. The digital revolution is generating a massive exposure of young people to a variety of online content, including video. This research aims to study the influence of an exposure to socio-scientific controversy presented by a set of videos on written and argued productions by 7th students. We analyze the impact of this exposure as well as activities related to the videos on the quality of the arguments produced. We show that subjects formulate richer answers and take more into account the arguments opposed to their position after consulting the videos, and that a sorting activity of the videos makes them produce more arguments in their answers.

Keywords. Video learning, Socio-scientific controversy, Argumentation, Counter argumentation

1 Introduction

Cette recherche s'inscrit dans le cadre du projet DIMEDD[1] et porte sur les compétences argumentatives mobilisées à l'écrit par des élèves de 5ème au sujet d'une controverse socio-scientifique présentée en vidéo.

1 Le projet DIMEDD (Dispositif Multimédia pour l'Education au Développement Durable) est un projet de recherche E-Fran (Espaces de formation, de recherche et d'animation numérique) porté par le Laboratoire Cognition Langues, Langage et Ergonomie (Université de Toulouse – Jean Jaurès & CNRS) et soutenu par le Ministère de l'Education Nationale et de la Jeunesse.

Les compétences argumentatives évoluent avec l'âge. Dès 4 ans, les enfants se montrent capables de justifier leur position à l'oral pour obtenir quelque chose (Weiss & Sachs, 1991) mais c'est au moins à partir de 10 ans qu'émergent des formes argumentatives plus complexes intégrant plusieurs arguments connectés entre eux (Coirier et Goldier, 1993), des concessions ou des raisonnements contre argumentatifs (Brassart, 1992 ; Kuhn, 2003). Il est donc pertinent d'étudier l'effet d'une exposition à une controverse auprès de pré-adolescents sur leur capacité à intégrer ou discuter des arguments opposés à leur position.

Du fait d'un usage de plus en plus fréquent de vidéos par les jeunes dans leur vie privée ou par les enseignants dans un contexte éducatif, ce format apparaît intéressant pour réaliser cette étude. La dimension visio-dynamique de ce format présente de nombreux avantages pédagogiques, mais la mobilisation différente des ressources attentionnelles qu'il implique par rapport à du contenu strictement textuel suggère également ses limites. Plusieurs recherches ont cependant montré l'intérêt de mettre en place des interventions qui impliquent l'activité de l'apprenant dans la consultation de vidéos pour rendre ce format d'autant plus efficace dans l'intégration de connaissances (Merkt, Weigand, Heier, Schwan, 2011 ; Salomé, 2018) Cette étude intègre donc deux types d'intervention dont l'impact sur les compétences argumentatives des élèves sera étudié. La première est la présentation d'une vidéo pédagogique sur l'argumentation et la seconde une activité de tri thématique à réaliser sur l'ensemble des séquences visionnées.

2 Méthode

Sujets. 753 élèves de 5ème (âge moyen : 12 ans) provenant de 6 établissements différents ont participé à cette étude.

Procédure. L'étude s'est déroulée en deux phases.

Pré test. Durant une première phase, les élèves étaient invités à répondre devant un ordinateur à la consigne suivante : « *Imagine la situation suivante : grâce à l'argent récolté avec une vente de gâteaux, ta classe décide d'organiser un repas pour fêter la fin de l'année. Tu es chargé, avec une camarade de classe - Mélissa -, de t'occuper d'acheter les aliments nécessaires pour cuisiner le menu. Mélissa te propose que vous achetiez les aliments pour cuisiner le menu dans leur version bio plutôt que des aliments conventionnels (non-bios). Qu'en penses-tu ? Serais-tu d'accord ou pas d'accord avec Mélissa pour acheter des aliments bios plutôt que conventionnels pour cuisiner le repas ? Réponds dans le cadre ci-dessous, en expliquant ta réponse et en faisant des phrases.* ». Une moitié des élèves était ensuite invitée à consulter une vidéo pédagogique de quelques minutes donnant des conseils pour enrichir ses compétences argumentatives.

Post test. Durant la seconde phase de l'étude, les élèves consultaient depuis une plateforme en ligne 18 courtes séquences vidéo (durée moyenne : 22s) apportant chacune un argument en faveur ou en défaveur de l'agriculture ou de l'alimentation

biologique (9 vidéos pour, 9 vidéos contre). Une moitié des élèves devait trier ces vidéos au fil du visionnage selon 3 thématiques (santé, environnement, autre) et selon la position qui était défendue (en faveur ou en défaveur de l'alimentation biologique). L'ensemble des élèves avait ensuite à répondre à nouveau par écrit à une question ouverte leur demandant comme en pré-test de choisir entre un menu biologique ou conventionnel et d'argumenter leur réponse.

Analyse. Le corpus constitué à partir des réponses aux deux questions ouvertes a fait l'objet d'une analyse manuelle à partir d'une grille élaborée pour cette recherche. La robustesse de cette grille a été évaluée par le codage parallèle d'une partie des données par une personne extérieure à l'étude. Le taux de correspondance entre les deux codages a été de 96%, seuls trois sous-critères secondaires ont obtenu un coefficient de kappa inférieur à 0,6.
Les analyses avaient pour objectifs de déterminer et comparer entre les deux phases le nombre de mots et d'arguments par réponse, le type de justifications (uniquement des arguments en faveur de sa position, ou présence d'arguments en faveur de la position opposée) et la présence ou non de raisonnements contre argumentatifs (critique d'un argument en faveur de la position opposée) ainsi que de plusieurs connecteurs argumentatifs dans les réponses.

3 Résultats

Les résultats montrent une évolution significative ($p < .001$) en nombre de mots et d'arguments dans les réponses après que les élèves ont vu les vidéos. Ceux-ci ont produit en moyenne des réponses de 20,9 mots avec 1,19 arguments en pré-test, et de 29,6 mots avec 1,77 arguments après avoir vu les vidéos. On peut voir dans le tableau ci-dessous que les élèves ont plus souvent apporté aucun ou un argument dans leur réponse en pré-test, et plus souvent deux, trois, ou quatre arguments dans leur réponse en post-test.

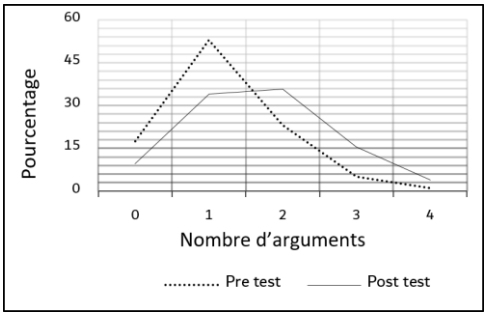


Fig. 1. Distribution des élèves en pourcentages selon le nombre d'arguments fournis dans leurs réponses avant (pré-test) et après (post-test) avoir vu les vidéos.

Les élèves ont également plus souvent pris en compte dans leurs réponses les arguments opposés à leur position après avoir vu les vidéos. On observe une augmentation des réponses intégrant des arguments des deux parties (28,9% vs. 11,4%, $p < .001$), ainsi que

des raisonnements contre argumentatifs (15,8% vs. 4,6%, p<.001). De plus, des relevés effectués sur différentes conjonctions exprimant une concession (*mais, même si, malgré, sauf, alors que, cependant…)* ont montré que l'expression « *même si* » avait été employée dans un plus grand nombre de réponses en post-test (13,3% vs. 2,9%, p <.001).

Des ANOVA à mesures répétées ont été effectuées pour évaluer l'effet de la vidéo pédagogique et du tri thématique sur les réponses des élèves. Nous n'avons relevé aucun effet de la vidéo pédagogique mais les élèves ayant eu à réaliser le tri thématique au fil de la consultation des vidéos ont produit ensuite plus d'arguments par rapport au pré-test [$F (2, 522) = 7,98$; $p = 0.005$; $\eta^2p = 0.012$].

4 Conclusion

Nous avons montré que l'exposition par vidéo à un ensemble d'arguments opposés sur un sujet de controverse (l'alimentation biologique) avait un effet positif sur les compétences argumentatives d'élèves de 5ème à l'écrit. Ceux-ci se sont montrés capables d'enrichir leurs connaissances sur le sujet et d'exprimer cet apprentissage en produisant des argumentations plus longues et plus riches. Leurs références plus nombreuses à des arguments opposés à leur position, ainsi que leurs usages plus fréquents de raisonnements contre argumentatifs ou de certains connecteurs verbaux indiquant une concession montrent qu'ils sont en mesure de relativiser leur position et de comprendre la multiplicité des points de vue sur un même sujet. Cet effet peut être interprété comme l'expression d'un point de vue relativiste (Kuhn, Cheney, Weinstock, 2000) dont l'émergence est caractéristique de la pré adolescence. Les résultats ont également montré l'utilité d'une activité de tri thématique des vidéos sur la richesse argumentative des réponses, suggérant l'intérêt d'exercices actifs et originaux en lien avec la consultation de vidéos pour améliorer leur impact et leur efficacité pédagogique.

Références

1. Brassart, D.G.: Negation, Concession and Refutation in Counterargumentative Composition by Pupils from 8 to 12. Argumentation, Vol. 6 (1992) 77-98
2. Coirier, P., Golder, C.: Writing Argumentative Text : a Developmental Study of the Acquisition of Supporting Structures. European Journal of Psychology of Education, Vol 2 (1993) 169-181
3. Cojean, S. : Etayage des Activités de Recherche d'Information et d'Apprentissage en Environnement Vidéo : Apports de la Segmentation et de la Structuration. Thèse de doctorat, Université de Bretagne Loire (2018)
4. Kuhn, D.: The Development of Argument Skills. Child Development, Vol. 74 (2003) 1245-1260
5. Kuhn, D., Cheney, R., Weinstock, M.: The Development of Epistemological Understanding. Cognitive Development, Vol. 15 (2000) 309-328
6. Merkt, M., Weigand, S., Heier, A., Schwan, S.: Learning with Video vs. Learning with Print : The Role of Interactive Features. Learning and Instruction, Vol. 21 (2011) 687-704

7. Weiss, D.M., Sachs, J.: Persuasive Strategies Used by Preschool Children. Discourse Processes, Vol.14 (1991) 55-72

3.3 Le numérique au service de la transmission des savoirs.

Anne Laure Bialès

Le numérique au service de la transmission des savoirs.

Par Anne Laure BIALES.
Université Paul Valéry Montpellier 3
Praxiling- CNRS

Résumé:

Le numérique fait partie du paysage éducatif, que ce soit dans les classes, dans les pratiques pédagogiques et dans les programmes. Les « technologies de l'information et de la communication pour l'éducation » ou TICE sont bel et bien devenues un objet d'enseignement. Mais, le numérique peut également être utilisé comme objet médium aux apprentissages. Si, ces dernières années les formations hybrides comme les classes inversées se sont multipliées, notamment dans l'enseignement supérieur, elles ont encore besoin de reconnaissance et de légitimité. Un élève ayant suivi un enseignement à distance peut-il se prévaloir des mêmes compétences validées par ses camarades en présentiel ? Que dire quand le numérique est l'unique choix possible pour dispenser un enseignement ?

Mots clés: TICE - enseignement-apprentissage - classe inversée.

Abstract:

Digital is part of the educational landscape, in classrooms, teaching practices and programs.

The "information and communication technologies for education" or TICE have indeed become a subject of instruction. But digital can also be used as a learning medium. If, in recent years, hybrid formations such as flipped classroom have multiplied, especially in higher education, they still need recognition and legitimacy. Can a student who has followed a distance education claim the same skills validated by his / her classmates? What to say when digital is the only possible choice for teaching?

Key words: TICE- teaching - learning - flipped classroom

I) Le numérique, les élèves, les enseignants.

Le numérique est de plus en plus présent dans le quotidien des élèves. Leurs compétences en informatique dépassent parfois celles de leurs formateurs. Si le numérique attire autant les élèves, c'est grâce à son caractère innovant, en constante évolution. Le numérique a cette merveilleuse faculté de se renouveler et de surprendre ses utilisateurs. Alors quand ce sont les plus jeunes qui découvrent ces nouvelles technologies, il est important que le formateur les guide. Car au-delà des compétences techniques que les élèves aiment développer, il y a le côté éthique dont ils n'ont pas toujours conscience. Il faut, pour reprendre les termes

de Karsenti : « Amener les apprenants à agir en citoyens numériques ».[1] En tant que formateurs, nous devons restés objectifs pour offrir à nos élèves un enseignement adapté et fonctionnel.

II) La classe inversée : définition

Ces dernières années, les élèves ont pu expérimenter de nouvelles méthodes d'enseignement qui allient transmissions des savoirs et numérique. Si les résultats des divers projets sont toujours à prendre dans leur contexte, il est important d'en faire ressurgir les bons comme les mauvais côtés. Le développement d'internet a largement permis la diffusion de cours enregistrés, dans un premier temps, puis sous forme de fichiers et de nos jours, de capsules vidéo. La classe inversée provient des États-Unis sous l'appellation de "flipped classroom". Ce concept est venu d'un constat fait par deux enseignants, Bergman et Sams. [2] Les premières classes inversées ont vu le jour dans les années 90. Les devoirs à la maison étaient une pratique courante. Mais quand un apprenant a des difficultés d'apprentissage, faire ses devoirs, parfois sans aide, ne fait que renforcer (voire empirer) ses difficultés. Les deux enseignants avaient également remarqué que l'absentéisme nuisait aux apprentissages. Ils voulaient alors pouvoir offrir leurs cours aux élèves absents via une vidéo. Avec la mise en ligne de ses cours, Sams s'est rendu compte que les élèves absents posaient

davantage de questions lors du cours suivant, que ces élèves étaient plus impliqués dans son cours que ceux qui l'avaient suivi "en direct". C'est « une approche pédagogique consistant à inverser et à adapter les activités d'apprentissage traditionnellement proposées aux étudiantes et étudiants en utilisant en alternance la formation à distance et la formation en classe pour prendre avantage des forces de chacune. » [3] La classe inversée s'oppose au modèle traditionnel. Avec la classe inversée, les élèves avancent à leur rythme et l'enseignant cible plus facilement les besoins et les demandes de chacun. Cela semble être un système parfait si on veut faire de la différenciation dans sa classe, et pourtant... La classe inversée a ses limites. L'enseignant a moins de contacts avec ses élèves, la question de son statut et des relations qu'il entretient avec eux peut venir parfois lui donner envie de revenir au modèle traditionnel. Les élèves, très au fait et très attirés par le numérique, pourraient prendre cela pour un jeu, de moins prendre au sérieux l'acquisition des savoirs et le développement de leurs compétences. Narcy-Combes (2005) remarque que les élèves sont davantage « isolés » quand ils sont face à leur ordinateur.

L'enseignant est plus libre de configurer sa classe, sa programmation et ses conduites d'apprentissage. Ces occasions multiples de faire de la différenciation qu'offre la classe inversée lui permet de faire des évaluations

plus formatives que sommatives. Les élèves collaborent quand ils sont en cours. Dans le cadre de la classe inversée, les élèves sont actifs, co-construisent le sens ensemble, contribuent activement au développement de leurs compétences. Ils en sont responsables. Il est aussi important d'avoir des temps réflexifs sur sa pratique, de connaitre les limites de la classe inversée et de ne pas croire que l'enseignant est remplacé et remplaçable. La classe inversée redonne à la classe son rôle premier : la construction et la co-construction des apprentissages, des savoirs et des savoir-faire. Clark et Leonard disaient déjà, en 1985 : « Les ordinateurs ne contribuent guère plus à l'apprentissage que le camion qui fournit les victuailles aux magasins ne peut améliorer l'alimentation d'une communauté. Acheter un camion n'améliore pas plus la qualité de l'alimentation qu'acheter un ordinateur n'améliore l'accomplissement de l'étudiant. La qualité de l'alimentation provient d'une bonne adéquation entre les aliments fournis et les besoins des personnes. De manière comparable, la qualité de l'apprentissage provient d'une balance correcte entre les méthodes d'enseignement et les besoins des étudiants. » [4] Avec l'hybridation des formations, l'apprenant apprend à apprendre, toujours aidé par l'enseignant.

III) Le numérique à l'école : les TICE et leurs enseignements

Enseigner les TICE à l'Ecole va bien au-delà de l'utilisation de l'ordinateur et du traitement de texte. C'est savoir utiliser l'outil « ordinateur » et l'outil « internet ». Il y a de ce fait toute une éducation au numérique qui est nécessaire pour que nos élèves vivent à l'heure du numérique de façon sereine. Selon Hoechsmann et DeWaard (2015), il est nécessaire d'enseigner la « littératie numérique » [5], concept difficile à définir. On parle aussi plus largement de « culture numérique ». Pour Gauchet : « L'introduction du numérique entraîne une transformation profonde des manières d'enseigner et plus généralement du rapport de la société à l'enseignement. » [6] Les enseignants restent ceux qui transmettent les savoirs. Or, ils sont confrontés chaque jour à la concurrence avec internet. Les connaissances sont disponibles, avec plus ou moins de fiabilité. Comment rester légitimes dans un monde où la connaissance est en libre accès ?

IV) Dispenser un enseignement grâce au numérique.

Le projet est né des besoins du terrain. Les apprenants allophones (les EANA : élèves allophones nouvellement arrivés) sont des élèves qui ont pour langue maternelle une autre langue que le français. Afin qu'ils

soient le plus rapidement inclus dans les classes dites « ordinaires », ils bénéficient d'un enseignement spécialisé dispensé par des enseignants spécialisés. Quelques heures par semaine, ils sont en UPE2A (unité pédagogique pour élèves allophones arrivants). Mais, les enseignants d'UPE2A sont peu nombreux. De plus, selon les établissements scolaires, le nombre d'élèves allophones ne permet pas d'ouvrir un tel dispositif. Dans le cadre de notre recherche sur « la littérature via les TICE dans l'enseignement du FLS », nous avons pu dispenser des cours de français langue seconde (FLS) aux élèves allophones qui n'avaient pas d'enseignants spécialisés. Nous avons choisi la plateforme « Edmodo ». Grâce à l'enseignement via l'ordinateur ou la tablette, les élèves pouvaient rester en classe et travailler sur des contenus dédiés à leurs besoins. De plus, les enseignants, qui devaient inclure les élèves allophones dans leurs classes, pouvaient y trouver des ressources pédagogiques afin de les sensibiliser aux besoins des EANA. Ils pouvaient ainsi réinvestir le travail produit par les élèves en ligne. Si certes, pour les besoins de notre travail de recherche, nous avons mis un fort accent sur la littérature, les élèves pouvaient également trouver des activités autour du lexique, de la grammaire, des activités culturelles et ludiques… Toutefois, pour Narcy-Combes qui cite Brodin (2004) mettre tout à disposition en ligne « ne signifie pas que les intéressés s'en

serviront, et qu'il y a toujours un écart considérable entre ce qui est prescrit et ce qui est réellement accompli, aussi bien pour les enseignants que pour les apprenants. » [7] Il ne faut pas se contenter de transmettre et recevoir du savoir.

V) Du côté des enseignants.

Dans le cadre de notre recherche, nous avons demandé aux professeurs qui suivaient les élèves utilisateurs de la plateforme et qui le souhaitaient, de répondre à un questionnaire sur l'utilisation des TICE dans le cadre de l'enseignement. Tous déplorent tout d'abord un manque de formation aux nouvelles technologies. Tous sont également conscients que la génération d'élèves aujourd'hui possède des compétences qui dépassent les leurs. Comment faire face à des élèves experts quand on est censé l'être ? Le numérique est perçu comme contraignant pour ceux qui n'en sont pas maîtres. Malgré tout, les discours de ces enseignants se sont avérés positifs car ils avaient besoin de cette plateforme pour inclure au mieux les élèves allophones dans leurs classes. Le numérique ne s'est pas imposé à eux de manière brutale et frontale. Il s'est avéré être utile pour transmettre des savoirs, qu'ils ne possédaient pas, à leurs élèves.

V) Conclusion.

Dans le cas de notre projet autour de la dispense de cours de FLS via une plateforme numérique, la légitimité des savoirs transmis n'a jamais été remise en question. Le numérique peut servir à la transmission des savoirs dans un cadre contrôlé par l'expert, l'enseignant. Il est nécessaire d'éduquer les élèves au numérique pour qu'ils en aient une maitrise réfléchie et distanciée. Des règles sont à poser et des adaptations sont à apporter en fonction des besoins, des compétences mais aussi en fonction des contextes sociaux des établissements scolaires. Le numérique ne peut en aucun cas participer au creuset des inégalités entre les élèves. Les savoirs transmis via le numérique doivent être accessibles à tous.

Références

[1] Karsenti T., 21 Idées pédagogiques pour préparer et engager les apprenants avec le numérique, (2019).

[2] La classe inversée : que peut-elle apporter aux enseignants ?,

https://www.reseau-canope.fr/agence-des-usages/la-classe-inversee-que-peut-elle-apporter-auxenseignants.html, (2015), consulté le 04 février 2017

[3] « Faire la classe mais à l'envers : la flipped classroom », www.usherbrooke.ca/ssf/veille/bulle;ns/2011-•-2012/novembre-•-2011/le-•-ssf-•-veille/ faire-•-la-•-classe-•-mais-•-a-•-lenvers-•-la-•-flipped-•-classroom, (2011), consulté le 05 mars 2016

[4] Clark R.E et Leonard S., Computer Research Confounding , The annual meeting of the American Educational Research Association, Chicago, Illinois, (1985).

[5] Hoechsmann M. et DeWaard H., Définir la littératie numérique,

https://eduscol.education.fr/numerique/tout-le-numerique/veille-education-numerique/archives/2015/mai-2015/definir-la-litteratie-numerique, (2015), consulté le 10 mars 2018

[6] Davidenfoff E. et Gauchet M., Internet bouleverse-t-il vraiment l'éducation ?, dans l'Express, (2014).

[7] Narcy-Combes, J-P., Didactique des langues et TIC: Vers une recherche-action responsable, Ophrys Éditions, (2005).

3.4 Instrumentalisation de l'intelligence collective ?

Pierre Bellet

Colloque Jeunes Chercheurs PRAXILING 2019

Instrumentalisation de l'intelligence collective ?

Pierre Bellet
LHUMAIN, Montpellier, France
pierre.bellet@humanconnect.eu

Résumé. Entre les évolutions et révolutions liées aux nouvelles technologies, de nouvelles formes de littératie numérique émergent en permanence. De l'actuel au virtuel, de l'écrit à la donnée, cette réflexion porte sur la mise en œuvre d'une intelligence collective au service des nouvelles formes d'apprentissages, ainsi que de leurs nouveaux objets et nouveaux besoins. Au travers de la littérature, nous sélectionnons quelques axes théoriques pour orienter la démarche de recherche : l'individu en tant qu'apprenant, le sens de l'intelligence, et l'instrumentalisation, le détournement des outils à des fins d'apprentissage par la co-conception et la résolution de problèmes. Cette sélection permet la problématisation de plusieurs dimensions incertaines, nécessitant l'adaptation d'une méthodologie pour parvenir à des résultats cohérents avec les questions et intentions de recherches. Plusieurs pistes méthodologiques sont ainsi présentées, rassemblant un ensemble de méthodes et outils alliant pédagogie et numérique. Ce papier présente un positionnement et des intentions. Il se termine sur des perspectives, sans conclusions, nous invitant naturellement aux discussions.

Mots-clés. Intelligence Collective, Instrumentalisation, Technologie numériques, Pédagogie

Abstract. Between the evolutions and revolutions linked to new technologies, new forms of digital literacy are constantly emerging. From the actual to the virtual, from the written language to the data, this reflection focuses on the implementation of a collective intelligence at the service of new forms of learning, as well as their new objects and new needs. Through literature, we select a few theoretical axes to guide the research process: the individual as a learner, the meaning of intelligence, and instrumentalization, the use of tools for learning purposes through co-design and problem-solving. This selection allows the problematization of several uncertain dimensions, requiring the adaptation of a methodology to achieve results consistent with the research questions and intentions. Several methodological approaches are presented, bringing together a set of methods and tools combining pedagogy and digital technology. This paper presents a positioning and intentions. It ends with perspectives, without conclusions, naturally inviting us to the discussions.

Keywords. Collective Intelligence, Instrumentalisation, Digital Technology, Pedagogy

1. Introduction

Si les outils que nous avons su développer nous ont mené vers la maîtrise progressive de l'ordre des choses matérielles à plus ou moins large échelle, nos dernières technologies nous ont également propulsé vers l'exploration de l'ordre des choses immatérielles, abstraites, virtuelles, difficilement délimitables ou mesurables. Nous pouvons ainsi nous représenter avec un certain renouveau de technicité, des choses allant de l'infiniment petit à l'infiniment grand, de l'abstrait au concret, du *virtuel* à l'*actuel* (Deleuze, Parnet, 1996 ; Lévy, 1995 ; Maïlat, 2008).

Dans le contexte de l'apprentissage humain, impliquant notamment les technologies numériques (Vinck, Clivaz, 2014), la littérature scientifique foisonne d'articles et d'expérimentations impliquant de nouvelles méthodes, instrumentations, littératies (Bourdeloie, 2012) et manières de concevoir ou d'expérimenter l'apprentissage. On peut les regrouper sous l'appellation de "formations innovantes", formes hybrides de virtualisation de l'apprentissage : *e-learning*, *EAD*, *FOAD*, *MOOC*, *xxOC*, *BYOD*, *serious game*, *deep learning* avec des indications de *versionning* héritées de la pensée informatique : *1.0*, *2.0*, *3.0*. Il existe ainsi une multitude d'exemples permettant de reproduire et d'adapter des expérimentations ayant pour objectif une ou plusieurs des nombreuses problématiques liées à l'apprentissage (motivation, efficacité, réussite, performance), avec parfois des limites, des mésusages ou des discussions qui animent les communautés scientifiques (Bocquet, Bruillard, Cornu, Guignolet, Moatti, Véran, 2014).

Nous nous intéressons ici plus spécifiquement à quelques uns de ces exemples, afin de développer un modèle d'intention pour favoriser l'émergence d'une forme d'intelligence collective dans des contextes d'apprentissage en groupe.

2. Littérature et axes théoriques

Mais avant même de parler du groupe ou du collectif, il s'agit d'abord de réussir à définir l'individu. Si l'individu est souvent mis au cœur du processus (en tant qu'apprenant, en tant qu'utilisateur, en tant qu'acteur), il n'en est pas simplement définissable pour autant, se démultipliant en variables dans le temps, l'espace, le contexte, etc. Comment donc l'inclure dans son intégrité et sa subjectivité au travers d'une temporalité et un espace défini, sans le réduire à une évaluation trop incertaine qui remet en question sa validité ou son objectivité ?

2.1. L'Individu, cet invariant variable

Parmi l'ensemble des variables individuelles, il en est dont on sait précisément l'importance, mais que l'on peut difficilement mesurer et intégrer dans un modèle, surtout en situation ordinaire d'apprentissage. Pour n'en citer que quelques-uns, les aspects socio-économiques, psychologiques, physiologiques ou cognitifs sont autant d'ensemble de variables complexes que l'on ne semble pouvoir qu'effleurer dans leurs mesures ou exploitation à partir des résultats.

Si l'apprenant est un individu en construction, il s'agit de définir, en plus des variables qui le définisse à un instant *t*, ce qui est moteur de l'apprentissage, ce qui

mobilise l'effort et mène à un autre instant. Les travaux de Bandura (1986), sur la théorie sociocognitive, ou ses concepts de sentiments d'efficacité personnelle et collective souvent mis en avant, permettent de constater des effets sur une meilleure autorégulation des apprentissages, un renforcement de la volition, et une amélioration de la performance (Galand et Vandele, 2010). La motivation est également un facteur déterminant de la réussite (Bandura, 2000 ; Houart, 2017). Certaines dimensions psychologiques peuvent ainsi faire l'objet de leviers de transformation par une instrumentation individuelle et conscientisée, dans le but de rendre l'apprenant plus autonome et plus efficace (Mailles-Viard Metz, 2015).

2.2. Guère d'intelligences ?

Prise dans les réseaux du langage, on peut difficilement trouver un consensus sur une définition de l'intelligence. Elle est aussi multiple que toutes les disciplines qui s'y intéressent, et tous les domaines auxquels elle s'applique.

On peut alors tâché de définir l'intelligence comme la capacité à comprendre, apprendre, par un procédé de traitement entre information et action afin de s'adapter à de nouvelles situations. Son étymologie latine nous renvoie aux notions de « faculté de percevoir, compréhension », ou encore actions de « discerner, saisir, comprendre »). Composé du préfixe *inter-* (« entre ») et du verbe *lĕgĕre* (« cueillir, choisir, lire »), il y a originellement cette idée d'un choix, d'une sélection, de relier par la compréhension et le raisonnement.

Intéressons-nous maintenant à trois typologies possibles de celle-ci : l'intelligence générale (IG), l'intelligence artificielle (IA) et l'intelligence collective (IC).

L'IG et son discuté « facteur g », correspond aux modèles les plus avancés dont nous disposons pour étudier ses variables, notamment grâces à des tests psychométriques (e.g., Weschler, Stanford-Binet) qui permettent d'évaluer le tout aussi discuté « Quotient Intellectuel ». Les travaux du psychologue McGrew (1997) ont donné lieu au modèle Cattell-Horn-Carroll (CHC), adapté depuis continuellement jusqu'à ce jour, avec les récents apports neuropsychologiques et psychométriques. Si le modèle original inclut différentes strates avec des facteurs moteurs, perceptifs, attentionnels, de gestion de la connaissance et de vitesse, les récents travaux incluent notamment la dimension d'intelligence émotionnelle, marquant l'ouverture vers d'autres variables essentielles et nous renseignant peut-être, sur l'autre moitié de la variance dans les résultats de tests de QI (Deary, Penke, Johnson, 2010).

L'IA est citée à titre d'analogie, en ce qu'elle a d'intéressant à mettre en application par des théories et techniques ce que l'homme projette de l'intelligence humaine, avec pour objectif de simuler des intelligences au travers de machines. De théories en "artisanat de l'intelligence", nous industrialisons maintenant celle-ci au travers des machines, comme nous l'avons fait dans l'industrie automobile, durant la Révolution industrielle. Si l'on fait référence à *La Guerre des Intelligences* (Alexandre, 2017), sous-titrée "Comment l'Intelligence Artificielle va révolutionner l'éducation", nous sommes actuellement dans la phase 2 du développement de l'IA, c'est à dire la phase où l'on ne programme plus les IA, mais on les éduque, comme pour les humains, avec des logiques de *deep learning* ou *machine learning*. Le paradoxe soulevé ici étant qu'il est beaucoup plus facile de mesurer précisément des notions telles que la performance, l'efficience ou encore des mécaniques d'acquisition ou de traitement lorsqu'on les observe sur des machines programmées ; bien qu'elles soient relativement autonomes, elles restent encore systémiquement prévisibles puisqu'objets de nos conceptions.

L'IC peut être définie comme une « *intelligence partout distribuée, sans cesse valorisée, coordonnée en temps réel, qui aboutit à une mobilisation effective des*

compétences » (Lévy, 1994, p. 29). D'un point de vue linguistique, elle pourrait être analysée selon sa sémantique, grâce à un métalangage tel que l'IEML (Lévy, 2019), encore en développement expérimental mais dont le but est de permettre la représentation de celle-ci. Cette définition permet de garder distance avec une approche qui se concentrerait d'avantages sur des résultats ou des mesures rapportées à des normes. Il est question de genèse plus que de ses projections ou applications potentielles. Nous nous intéressons à l'IC comme un vecteur enseignable, capable de répondre aux problématiques complexes de nos sociétés humaines, nécessitant une contribution supérieure à celles d'individus isolés dans leur action. Il s'agit cependant d'aborder l'IC selon une co-construction à partir de petits groupes d'individus, et non de masses ou de foules dans leur ensemble.

2.3. Instrumentalisation : sujet(s), objet(s), activité(s), action(s)

La perspective instrumentale au sens de Rabardel (1995), permet de resituer les relations entre l'humain, les artefacts et les activités dans un cadre théorique. Au travers du processus de *genèse instrumentale*, il est question de deux orientations observables : l'*instrumentation*, portant sur le sujet et son appropriation de l'artefact par assimilation et accommodation ; et l'*instrumentalisation*, portant sur l'artefact "exproprié" par le sujet qui le détourne et l'adapte à ses besoins (Folcher, Rabardel, 2004).

Dans le cadre d'un apprentissage de type conception participative (Reggers, Khamidoullina, Zeiliger, 2002), avec par exemple, des modalités collaboratives et à distance requérant l'usage de médiation numérique, l'on conçoit usuellement un scénario pédagogique construit, permettant d'articuler l'acquisition progressive des savoirs et compétences à développer selon des objectifs pré-définis (Patrick, Huot, 2017). À titre d'exercice, l'on présente généralement un problème, plus ou moins défini (ou "*mal défini"* comme le rappelle Darses, Détienne et Visser (2001), relativement aux activités de co-conception) nécessitant un effort, une transformation et une forme de créativité pour parvenir à sa résolution (Fischer, Giaccardi, Eden, Sugimoto, Ye, 2005). Les individus en situation d'apprentissage doivent ensuite concevoir et produire une solution, au travers d'usage et de création d'artefacts qui, selon les modalités, peuvent leur permettre de mettre en œuvre différentes stratégies d'actions telles que la *coopération* ou la *collaboration*. Ainsi, l'objet même de la pédagogie numérique ou de l'ingénierie techno-pédagogique (Charnet, 2019) peut permettre de relier des usages (utilité, besoin) et des pratiques (*praxis*, action humaine) dans une visée instrumentalisée de l'apprentissage. Son évaluation peut ou doit être aussi ergonomique selon les technologies utilisées (Loup-Escande, Burkhardt, Richir, 2013).

L'instrument est de plus capable de produire des traces, que l'on peut convertir en données construites pour analyses et résultats induisant ou non des conclusions selon les hypothèses initiales. Mais il transforme aussi la pratique, de manière directe ou indirecte, il est objet médian, et a une influence autant sur les sujets, que les objets, produits ou résultats des activités observées. L'un des verrous scientifiques à lever concernerait cette mesure d'effets sans se perdre dans une complexité métrologique qui tend souvent à s'éloigner de l'humain, paradoxalement aux démarches engagées. L'une des pistes pourrait être ici, d'impliquer plus activement l'individu auto-observé dans l'analyse et l'utilisation des résultats d'un protocole expérimental dont il serait à la fois acteur et commanditaire (une démarche engagée et motivée par les sujets).

Dans le cadre de l'observation d'un phénomène tel que l'IC, impliquant l'analyse de l'ensemble des variables relatives à chaque individu dans leur ensemble interactionnel, avec une visée d'instrumentalisation pédagogique, il serait ensuite question de pouvoir

agréger ces données à un niveau *collectif-synchrone*, afin de les rendre activables pour favoriser la connaissance, le partage et l'efficacité des acteurs impliqués.

3. Problématisation et questions de recherches

À partir de ces principaux axes d'apports théoriques identifiés, plusieurs questionnements sont déjà soulevés.

Le premier concerne l'individu, ce qui le définit, son intégrité, sa personnalité, ses émotions, ses comportements, ses compétences et ses capacités à mobiliser éventuellement son intelligence personnelle au service du collectif. Il s'agit de questionner l'art et la manière de solliciter et d'accompagner un apprenant vers une certaine autonomie, afin de s'intégrer plus effectivement au sein de groupes opérant sur la résolution de problèmes plus complexes. Une bonne connaissance de soi et un apprentissage à l'ouverture aux autres, eux-mêmes pris dans cette dynamique de connaissance de soi, peuvent-il favoriser les effets de synergies et stigmergies collectives ? Doit-on considérer des effets de "compensation" ou de complémentarité dans la somme des différences interindividuelles ?

Le second concerne la définition même de l'intelligence, sa mesure, et sa transposition appliquée à un niveau collectif. Est-il ici question d'utiliser des échelles, des états de constructions, de états de production ou de produit fini ? L'évaluation de facteurs tels que ceux du CHC peut-elle être transposable ou corrélable à des effets issus de l'IC ? Doit-on rapporter l'IC à une mesure comparable (Green, 2015), en terme de gain de temps et de performances sur un résultat attendu et des configurations compétitives différentes (individu seul *vs* groupe non organisé *vs* groupe coordonné *vs* "IA") ?

Le troisième concerne davantage la dimension méthodologique et les différentes modalités pédagogiques, techniques, technologiques et pratiques pour favoriser la mise en œuvre d'une IC, qui puisse être synchrone et distribuée (ubiquitaire), sans cesse valorisée (impliquant une mesure ou une évaluation de la valeur), et mobilisant effectivement les compétences liées à la résolution d'un problème plus ou moins défini. Il y a ici un questionnement sur l'intégration et la collaboration participative des différents acteurs dans le processus de *Design Based Research* et de *Conception Centrée Utilisateur*. Il concerne autant les différents rôles que la valorisation permanente des différentes contributions dans un processus global porteur de sens pour chaque acteur impliqué.

4. Méthodologies

La mise en transversalité des disciplines concernées dans un tel contexte n'est pas évidente, mais fortement encouragée si l'on observe les plus de 25% de projets de l'ANR proposés sur des "*enjeux transversaux*", situés *à la croisée de plusieurs* secteurs scientifiques, comme les *sciences du numériques* ou les *SHS*. Bien que régulièrement appuyée par des recherches empiriques faisant de l'expérience sensible un vecteur d'accumulation d'observations et de faits mesurables, cette approche qui consiste à aller du concret à l'abstrait par induction ne parvient pas toujours à saisir suffisamment précisément l'intégralité de ses objets. Et lorsqu'il s'agit d'étudier des objets de recherche qui impliquent des sujets humains dans des situations complexes

généralement inversées (de l'abstrait au concret, i.e, dans une démarche de conception), il semble que certains paradoxes méthodologiques soient soulevés.

La communauté EIAH (Environnement Informatique pour l'Apprentissage Humain), qui pratique de fait une forme de transdisciplinarité, a su créer et partager des formes méthodologiques permettant de mieux cadrer ces variables contextuelles dans la mise en place d'expérimentation, la collecte de données, ainsi que leur évaluation. La méthode THEDRE (Mandran, 2017) en est un exemple, basée sur une démarche de qualité (i.e., *roue de Deming*, *PDCA*) relativement à l'environnement informatique et une approche empirique qui permet aux chercheurs d'adapter la méthode à leurs propres questions de recherche et ainsi de développer leur propre méthodologie.

Du chercheur, chercheur-enseignant au *technologue de l'éducation* (Denis, 2007) ou encore au *méthodologue* (Mandran, 2017), on constate que ces évolutions méthodologiques nous amènent au développement des mêmes compétences transversales que celles à enseigner aux individus apprenant.e.s, avec cette même dynamique de participation collaborative à des projets aux enjeux transversaux.

La démarche de construction d'un scénario pédagogique accompagnant l'émergence d'une forme d'IC réflexive et appliquée (Lévy, 2015), se base donc sur l'ensemble de ces éléments prenant, à ce jour, l'esquisse de la formalisation suivante :

- une *temporalité* sur le moyen ou long terme, permettant la *co-construction* d'un *apprentissage* et son *étude longitudinale*
- la définition d'une *tâche* suffisamment complexe, stimulante et créative, mais surtout *atteignable collectivement* (Bandura, 2000) sur un plus ou moins long terme, nécessitant alors une *organisation du travail* et des *instruments* pour le faciliter
- la caractérisation d'une *activité de co-conception*, notamment sur les *aspects cognitifs* se rapprochant de l'*évaluation de l'intelligence*, telle que développé dans le travail de Ruiz-Dominguez (2005), sans en faire une mesure compétitive
- la *co-conception* (*Conception Centrée Utilisateur*, *Design Based Research* ; Caelen, 2009 ; Heiwy, 2018) d'un *environnement informatique* avec les usagers, prenant en compte l'utilité et l'utilisabilité d'outils, plate-formes ou logiciels adaptés aux besoins et problématiques de la tâche
- l'accompagnement des individus dans la *connaissance de soi*, la *connaissance des autres*, et la complémentarité d'un collectif agissant vers un *but commun*
- l'accompagnement des individus sur l'*appropriation*, l'*instrumentation* et l'*instrumentalisation* d'outils numériques (e.g., communication, gestion, collaboratifs, synchrones) permettant de résoudre des problèmes au fur et à mesure de leur apparition
- des mécanismes d'*auto* et de *co-évaluation* permanentes et synchrones entre des individus formant groupe pour accomplir une tâche commune
- une *évaluation de la qualité* complémentaire à des *mesures quantitatives* non intrusives (ne venant pas perturber anormalement le déroulement de l'activité principale, particulièrement en contexte d'apprentissage)

5. Perspectives

Pour synthétiser ces besoins dans une plateforme opérationnelle aux spécifications technico-fonctionnelles adaptées, l'étude d'*outils de gestion*, de *travail collaboratif*, de *services de réseau social* et d'*apprentissage* se poursuit dans cette voie pour établir un cahier des charges et définir une solution logicielle. La formalisation d'un scénario

pédagogique et d'une tâche pour soutenir l'apprentissage et l'émergence de l'IC s'effectue en parallèle.

La poursuite de ces travaux propose de formaliser une invitation à dépasser les usages et pratiques des outils que nous utilisons pour observer, mesurer, évaluer les apprentissages, en les partageant directement et méthodologiquement avec les acteurs concernés. L'application expérimentale prendrait la forme d'un réseau social dédié à l'émergence d'IC co-construites, appliquées à des buts et objectifs d'apprentissage. Elle pourrait être aussi l'occasion de considérer les croyances, comme celles qui peuvent aspirer à l'efficacité, tel un dépassement de soi supérieur, qui se construit au travers de la pratique, de l'expérience et du collectif.

Si l'on en croit les pires pronostics, l'IC humaine se porte aujourd'hui comme une alternative à certaines projections futures pessimistes qui pourraient tendre vers des dérives eugénistes ou la "*fusion*" / *interfaçage* de nos intelligences personnelles avec de multiples IA (Alexandre, 2017).

Ainsi, nous misons sur la formation de collectifs, au travers d'une instrumentalisation des apprentissages, à la recherche d'une valeur qu'il resterait encore à (re)définir dans le « sens commun ».

Références

1. Alexandre, L. (2017). *La Guerre des Intelligences*. JC Lattès.
2. Bandura, A. (1986). *Social foundations of thought and action : a social cognitive theory*. Englewood Cliffs, NJ : Prentice Hall.
3. Bandura, A. (2000). Exercise of Human Agency Through Collective Efficacy. *Current Directions in Psychological Science*, *9*(3), 75–78. https://doi.org/10.1111/1467-8721.00064
4. Bocquet, F., Bruillard, É., Cornu, B., Guignolet, J., Moatti, D., & Véran, J.-P. (2014). Le numérique à l'école : évolution ou révolution pédagogique ? *Revue internationale d'éducation de Sèvres*, *67*, 103-118. https://doi.org/10.4000/ries.4129
5. Bourdeloie, H. (2012). L'appropriation des dispositifs d'écriture numérique: translittératie et capitaux culturel et social. *Études de communication*, *38*, 23-36. https://doi.org/10.4000/edc.3378
6. Caelen, J. (2009). Conception participative par «moments» : Une gestion collaborative. *Travail Humain*, *72*(1), 79-103. https://doi.org/10.3917/th.721.0079
7. Darses, F., Détienne, F., & Visser, W. (2001). Assister la conception: perspectives pour la psychologie cognitive ergonomique. *EPIQUE 2001, Actes des journées d'étude en psychologie ergonomique*, (1972), 11-20.
8. Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature reviews. Neuroscience*, *11*(3), 201-11. https://doi.org/10.1038/nrn2793
9. Denis, B. (2007). Articuler théories et pratiques en technologie de l'éducation. Dans B. Charlier (Éd.), *Transformation des regards sur la recherche en technologie de l'éducation* (pp. 31-52). Louvain-la-Neuve, Belgique : De Boeck Supérieur. https://doi.org/10.3917/dbu.charl.2007.01.0031
10. Deleuze, G., & Parnet, C. (1996). L'actuel et le virtuel. *Dialogues*, *250*(1995), 3-4.
11. Fischer, G., Giaccardi, E., Eden, H., Sugimoto, M., & Ye, Y. (2005). Beyond binary choices: Integrating individual and social creativity. *International Journal of Human Computer Studies*, *63*, 482-512. https://doi.org/10.1016/j.ijhcs.2005.04.014

12. Folcher, V., & Rabardel, P. (2004). Hommes, artefacts, activités : perspective instrumentale. Dans P. Falzon (Éd.), *Ergonomie* (pp. 251-268). Presses Universitaires de France « Hors collection ». https://doi.org/10.3917/puf.falzo.2004.01.0251
13. Green, B. (2015). Testing and Quantifying Collective Intelligence. *Collective Intelligence Conference*, 1-4.
14. Heiwy, V. (2018). Méthodes agiles, conception centrée utilisateurs : hybridation gagnante pour les projets innovants et pluridisciplinaires? Une application au cas du développement d'objets connectés. Dans *ERGO'IA 2018*. Bidart, France.
15. Houart, M. (2017). L'apprentissage autorégulé : quand la métacognition orchestre motivation, volition et cognition. *Revue internationale de pédagogie de l'enseignement supérieur*, *33*(2), 1-23.
16. Levy, P. (2015). Collective Intelligence for Educators. *Educational Philosophy and Theory*, *47*(8), 749-754. https://doi.org/10.1080/00131857.2015.1053734
17. Lévy, P. (2019). IEML : Le métalangage de l'économie de l'information. *Livre Blanc*. https://doi.org/10.13140/RG.2.2.11232.33281
18. Lévy, P. (1995). Sur les chemins du virtuel. *La découverte Poche*.
19. Loup-Escande, É., Burkhardt, J.-M., & Richir, S. (2013). Anticiper et évaluer l'utilité dans la conception ergonomique des technologies émergentes : une revue. *Le travail humain*, *76*(1), 27-55. https://doi.org/10.3917/th.761.0027
20. Mailles-Viard Metz, S. (2015). *Autonomie et apprentissage universitaire : aides et outils*. ISTE editions.
21. Mandran, N. (2017). THEDRE : langage et méthode de conduite de la recherche Traceable Human Experiment Design.
22. Patrick, P., & Huot, A. (2017). *Construire l'expertise pédagogique et curriculaire en enseignement supérieur: connaissances, compétences et expériences*. Québec : Presses de l'Université du Québec.
23. Rabardel, P. (1995). *Les hommes et les technologies; approche cognitive des instruments contemporains*. Paris : Armand Colin.
24. Reggers, T., Khamidoullina, I., & Zeiliger, R. (2002). Une « conception participative » centrée utilisateur. Dans D. Peraya (Éd.), *Technologie et innovation en pédagogie* (pp. 103-112). De Boeck Supérieur.
25. Ruiz-dominguez, G. A. (2005). *Caractérisation de l'activité de conception collaborative à distance : Etude des effets de synchronisation cognitive*. Institut National Polytechnique de Grenoble.
26. Vinck, D., & Clivaz, C. (2014). Les humanités délivrées. *Revue d'anthropologie des connaissances*, *8*(4), 681. https://doi.org/10.3917/rac.025.0681

4

Conclusions du colloque

Chrysta Pélissier, Arnaud Richard, Lucie Alidières-Dumonceaud (ITIC Praxiling UPVM3)(présentation)

Colloque Jeunes Chercheurs Praxiling 2019

PRAXILING

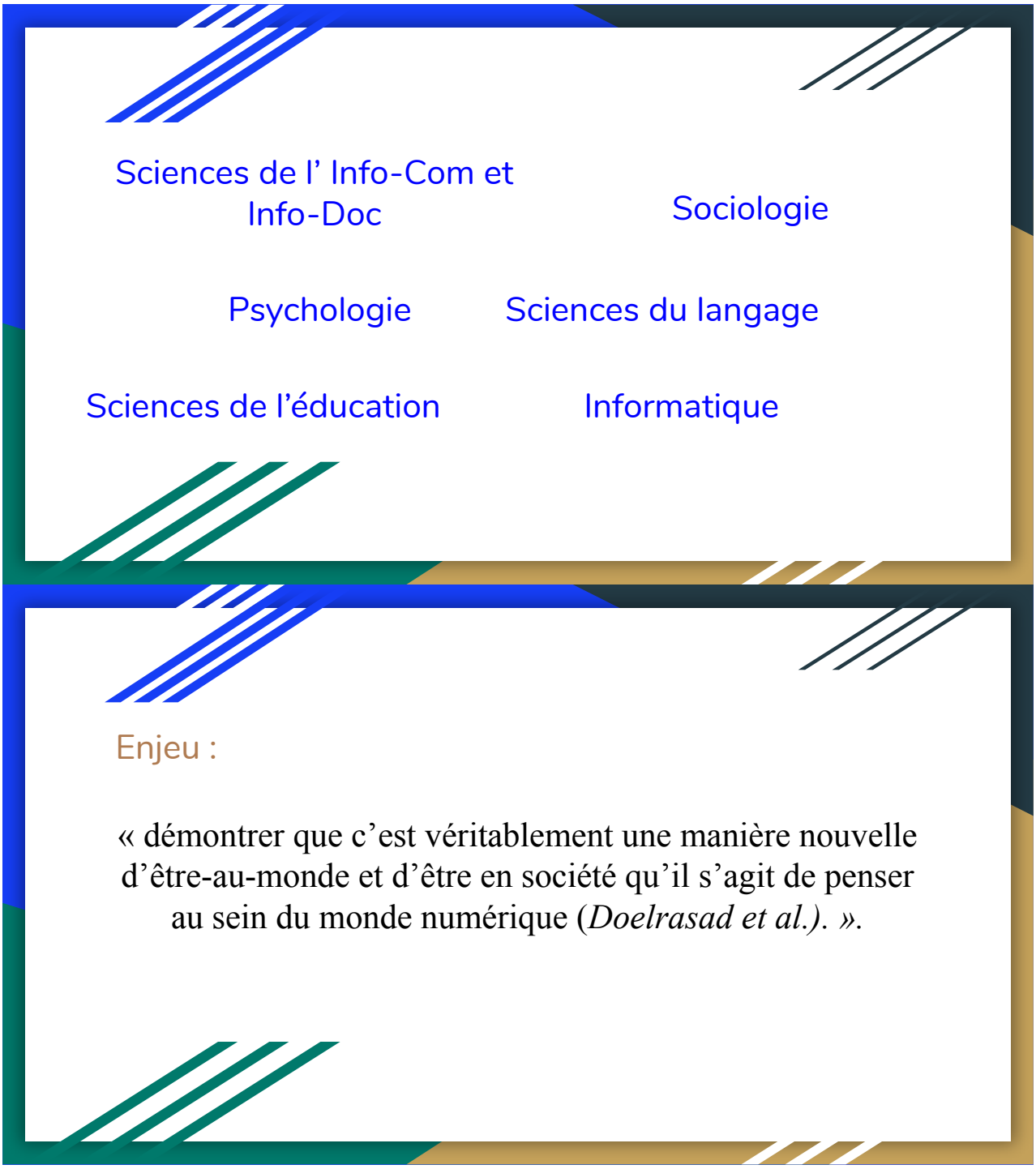
Représentation et transmission des connaissances à la lumière de l'innovation numérique

Clôture

Lucie Alidières-Dumonceaud, Chrysta Pélissier et Arnaud Richard

Contenu des communications :

- retours d'expériences
- réflexions sur la place des dispositifs/outils/ressources
- sur le rôle et la posture des enseignants
- représentations et imaginaires technologiques chez les étudiant-e-s
- modalités d'intégration, d'usage et d'apport de(s) outil(s)/ressource(s)/numérique(s)


Sciences de l' Info-Com et Info-Doc
Sociologie
Psychologie
Sciences du langage
Sciences de l'éducation
Informatique
Enjeu :
« démontrer que c'est véritablement une manière nouvelle d'être-au-monde et d'être en société qu'il s'agit de penser au sein du monde numérique (*Doelrasad et al.*). ».

Enjeu :

individu/collectif
autonomie

« démontrer que c'est véritablement une manière nouvelle d'être-au-monde et d'être en société qu'il s'agit de **penser** au sein du **monde numérique** (*Doelrasad et al.*). ».

avec
pour

- compétences
  -> compétences stratégiques / critiques et éthiques *(Guo)*
  -> adaptabilité / créativité / intelligence collective *(Doelrasad et al.)*
- appropriation d'outils numériques
  -> pour les enseignant-e-s *(Doelrasad et al.)*
  -> fracture numérique : celles et ceux qui ne savent pas / qui savent *(Guo)*
- prise en compte des parcours professionnels personnels
  -> développement personnel des doctorants *(Doelrasad et al.)*
  -> "l'image" dont l'enseignant-e est la productrice ou le producteur *(Ozil)*
- innovation numérique / pédagogique
  -> pour les enseignant-e-s
  -> pour les jeunes chercheuses et chercheurs

Moyens :

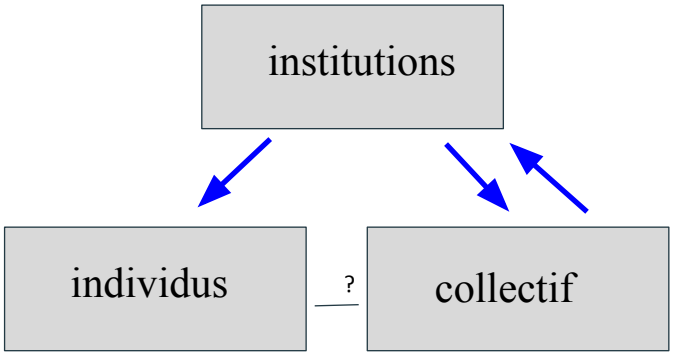
Solliciter les actrices et acteurs,
les former vers des changements notables,
dans une dynamique sociétale,
un environnement offrant des ouvertures

- ressources et activités
  -> données, conception, maîtrise *(Le Deuff)*

- prise en compte des parcours et images individuelles
  -> fluidité / adaptabilité en fonction des pratiques scientifique *(Doelrasad et al.)*
  -> diffusion d'une image personnelle *(Ozil)*

- de nouvelles situations
  -> loin d'une reproduction mécanique, prévisible mais plutôt des situations qui questionnent l'imprévisible *(Ozil)*
  -> outils méthodologiques de mesure / caractérisation de la confiance *(Samaniego)*
  -> des espaces d'apprentissages ludiques pour motiver et dynamiser les formations

Obstacles :



“échanger pour contourner, affronter ou envisager un changement….”

- politiques éducatives

-> “valeur-transmission” disparaît pour une “valeur-communication” (*Ozil*)

- solutions techniques à évaluer

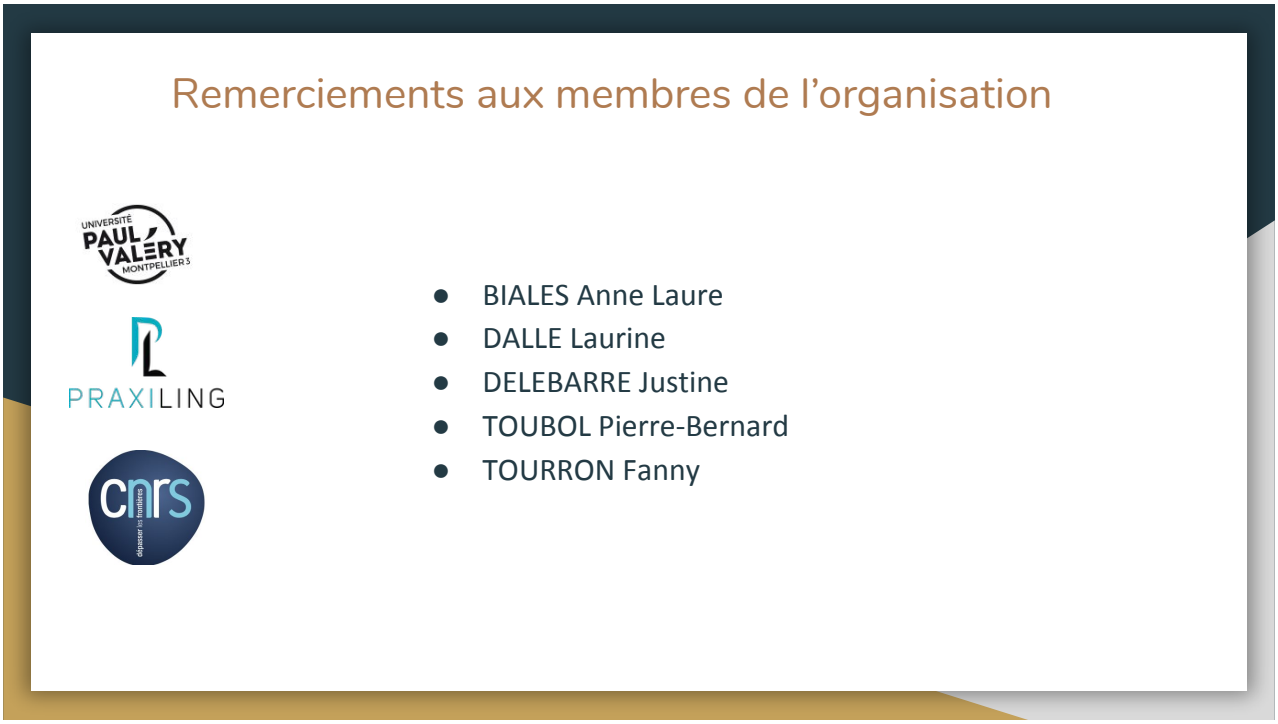
-> ne font pas toujours l’objet d’un processus d’appropriation (*Doelrasad et al.*)

-> effet de mode / télé-réalité / séries télévisées (*Ozil*)

- ressources à caractériser

-> manque d’outils de mesure de la validité de ces ressources (*Guo*)

-> confiance (*Samaniego*) / de peur / d’incertitude


Remerciements aux membres de l'organisation
UNIVERSITÉ PAUL VALÉRY MONTPELLIER 3
PRAXILING
cnrs
BIALES Anne Laure
DALLE Laurine
DELEBARRE Justine
TOUBOL Pierre-Bernard
TOURRON Fanny