

Encodage et analyse de données qualitatives

www.adjectif.net/spip/spip.php



Pour citer cet article :

Nicolas, Laura (2013). Encodage et analyse de données qualitatives
Apports et limites du logiciel Sonal comme outil de recherche en SHS. *Adjectif.net* Mis en ligne lundi 8 juillet 2013
[En ligne] <http://www.adjectif.net/spip/spip.php?article243>

Résumé :

Cette contribution présente le logiciel libre *Sonal*, outil d'analyse de verbalisations d'enseignants recueillies en entretiens d'autoconfrontation. Au-delà de la description du programme, on proposera une réflexion d'ordre épistémologique visant à circonscrire les apports d'un tel logiciel, en termes d'efficacité de traitement des données et d'ergonomie des interfaces. On relèvera cependant plusieurs éléments invitant le chercheur à une certaine prudence quant à l'interprétation des données traitées par cet outil informatique.

Mots clés :

Logiciels d'analyse, Logiciels libres



par **L. Nicolas**,

Doctorante en Sciences du Langage, Université Sorbonne Nouvelle – Paris III, DILTEC.

Didactique et TIC : entre pédagogie et recherche

La présente contribution s'inscrit dans le cadre d'une recherche de thèse en didactique du français langue étrangère (FLE), recherche qui vise à analyser le discours d'enseignants portant sur leurs pratiques de classe, et en particulier sur la manière dont ils réagissent aux interventions spontanées d'apprenants (ISA).

Avant d'aborder le traitement des données par *Sonal*, il importe de replacer l'utilisation de ce logiciel au sein du protocole méthodologique qui a été suivi, dans la mesure où ce protocole, issu d'une réflexion profondément transdisciplinaire, a presque intégralement été basé sur l'utilisation des TIC.

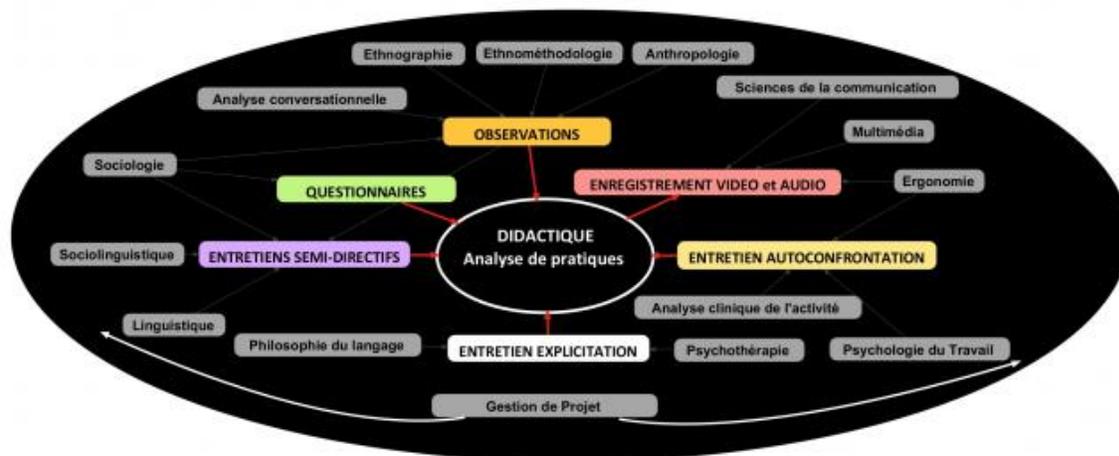


Schéma 1 : emprunts disciplinaires constitutifs du cadre méthodologique

L'analyse des interactions didactiques s'avère intrinsèquement liée, en termes théoriques autant que méthodologiques, à des disciplines connexes (Pothier, 2003 ; Beacco, 2011). Le schéma 1 souligne ainsi les emprunts effectués lors de la mise en place du recueil des données, à la sociologie, à l'ethnométhodologie et à l'analyse de l'activité, entre autres. L'utilisation de l'outil vidéo, en particulier, apparaît fondamentale dans cette étude. En parallèle de l'observation de classe et du recueil de données par questionnaires et entretiens semi-directifs, l'enregistrement sur support vidéo et audio a permis non seulement de garder une trace des activités [1] qui a ensuite servi à l'analyse des interactions, mais également de fournir un support aux commentaires d'enseignants réalisés après la classe, au cours d'entretiens d'autoconfrontation [2]. La pluralité des outils méthodologiques visait à fournir au chercheur des types d'observables pouvant être mis en comparaison : on a ainsi obtenu d'une part les données "factuelles" (interactions didactiques) qui ont été confrontées aux représentations des personnes interrogées via questionnaires et entretiens. Notre intérêt pour les apports du matériel audio et vidéo au cours du recueil des données s'est ensuite dirigé vers ceux des logiciels de traitement de ces données. Les verbalisations d'enseignants ont donc été entrées dans *Sonal* afin d'y être analysées.

Les apports de la synchronisation audiotextuelle : simplicité, rapidité, visibilité

Cette description du logiciel libre *Sonal* reste délibérément approximative puisqu'elle se veut une synthèse de celle qui a été réalisée par l'auteur du logiciel (Alber, 2010) [3]. *Sonal* est un récent logiciel d'archivage, d'encodage et d'analyse de matériaux qualitatifs enregistrés, qui ne peut être utilisé à ce jour que sous le système d'exploitation Windows [4]. Le premier avantage de ce logiciel nous semble résider dans la simplicité d'importation des données audio, vidéo et texte. Ayant déjà retranscrit les entretiens, nous n'avons rencontré aucun problème de compatibilité dans le processus d'importation des documents (.doc ou .rtf), contrairement à plusieurs autres logiciels testés en amont.

Une fois les fichiers importés, ils sont écoutés, visionnés, découpés et retranscrits selon le principe de la synchronisation audiotextuelle qui se rapproche fortement de ce qui se fait en montage vidéo. Le travail conjoint du texte et du son transforme la traditionnelle transcription de la bande son pour la faire évoluer en une tâche visant à *ajouter du son au texte*. Le texte est donc toujours replacé dans le continuum sonore de l'entretien, à l'intérieur des parties sélectionnées, ce qui permet (1) de revenir constamment à la bande son lors de l'analyse du texte, et (2) de retranscrire progressivement ou sélectivement ses corpus. Au-delà de l'aspect pratique d'une telle méthode, c'est le dépassement de la dualité oral-écrit inhérente à la retranscription qu'il s'agit ici de saluer : le son étant ancré au texte via la superposition des couches sonores et textuelles, ce n'est plus de la convention de transcription que dépend la bonne compréhension, par le lecteur-auditeur, de l'entretien, mais plutôt d'une adéquation rigoureuse entre ce qui est entendu et lu en même temps.

Une fois l'entretien découpé en autant de "chapitres" qu'il est jugé nécessaire par le chercheur, plusieurs possibilités

d'analyse lui sont offertes :

- **Analyse thématique** : via un système de coloration, chaque chapitre peut être thématiqué, selon des catégories auparavant définies par le chercheur. En ce qui concerne notre recherche, nous avons doublement "colorié" les entretiens (1) selon les différents thèmes abordés par l'enseignant (questions et commentaires des apprenants, planification du cours, gestion des tours de parole, motifs et buts de son action etc.) et (2) selon le mode discursif de ses verbalisations (interactif, théorique etc.) et les différentes "figures d'action" (action occurrence, d'expérience, canonique etc.) qui ont été définies par le modèle de Bulea & Bronckart (2012) et qui décrivent les modes d'expression de l'enseignant.
- **Analyse statistique** : à l'instar d'autres logiciels de traitement de données (*Modalisa*, *Lexico*, etc.), *Sonal* offre la possibilité d'analyser quantitativement une partie des données qualitatives. Toujours de manière simple et ergonomique, il est possible d'effectuer des analyses *lexicométrique* (filtrage des termes par différents indices et pourcentage, lemmatisation, entre autres), *chronométrique* (durée de parole, en fonction des types d'entretiens et des personnes interrogées) et *débitmétrique* (mots par minutes, en fonction des entretiens, des locuteurs et des thématiques). L'analyse lexicométrique s'est révélée particulièrement intéressante dans le cas qui nous occupe : on a tâché d'articuler le niveau lexicométrique avec celui des types de discours, en entendant certaines formules comme "je pense", "je crois", "il faut", comme indicatrices de modes discursifs relevant de l'ordre du jugement, de la croyance, de la norme etc.

Cette brève présentation du logiciel permet d'obtenir un aperçu de ce qui peut être effectué via cet outil, non seulement dans un but de recherche mais également dans un but de diffusion : la synchronisation audiotextuelle présente en effet l'avantage d'optimiser la visibilité de la présentation des résultats et des données lors de conférences, avec, par exemple, l'affichage automatique du texte par sous-titres sur la vidéo. En tant que logiciel libre et gratuit, *Sonal* est en amélioration constante, sur la base d'une participation collective de la part de la communauté scientifique.

Technicisation de la recherche : rôles du chercheur et des outils

Cette réflexion au sujet de l'utilisation de *Sonal* fait écho à celle qui a récemment été exprimée, ici même, par Komis, Depover & Karsenti (2013), qui visait à circonscrire le rôle des TIC dans une recherche qualitative, toutes disciplines confondues. Les auteurs mettent l'accent sur le paradoxe inhérent au fait que les TIC permettent à l'heure actuelle de pouvoir "quantifier" des données qui, de par leur nature qualitative, apparaissent justement comme des éléments "ne devant pas être quantifiés". Tenter un retour en arrière semble inapproprié, tant il reviendrait à renier l'apport considérable des TIC comme outils de recherche ; il apparaît cependant essentiel que chaque chercheur se pose la question de l'influence des outils qu'il a utilisés, de manière unique et située, dans une perspective théorique particulière et sur un terrain précis. Cette prise de recul sur notre propre expérience passée nous amène ici à formuler une modeste mise en garde au sujet de l'utilisation du logiciel *Sonal*, et, au-delà, des TIC comme outils de recherche. Il apparaît en effet essentiel d'entendre l'outil comme (1) constitutif du corpus de recherche, dans la mesure où le chercheur traite les données en fonction des possibilités du logiciel, de "ce qu'il lui donne à voir". Permettant d'étendre les possibles de la recherche, l'outil ne doit cependant pas être entendu comme (2) une alternative à la faculté de raisonnement et d'interprétation qui seuls sont l'apanage de la "personne-chercheuse" qui l'utilise.

Le logiciel, élément constitutif du corpus

Albero & Thibault (2009) ont montré l'hétérogénéité des approches théoriques qui sous-tendent l'utilisation des TIC en tant qu'outils pédagogiques et de recherche. L'approche anthropocentrée définie par Rabardel (1995), et les développements qu'elle a subi récemment (le "cours d'action" de Theureau et les pratiques de l'équipe d'Yves Clot dans le domaine de l'ergonomie, en particulier), permettent d'envisager l'utilisation, non seulement de la vidéo mais également des logiciels de traitement de données, dans la même lignée socio-constructiviste dans laquelle s'inscrit l'ensemble de notre étude. Le "réel" des interactions didactiques n'est pas entendu comme une réalité

indépendante des individus, mais plutôt comme un "construit" (Bachelard, 2000), issu de l'expérience subjective de chacun des participants. Dans cette approche, ce qui "fait" l'interaction didactique et ce qui révèle les enjeux qui la sous-tendent ne sont qu'en partie visibles sur les vidéos de classe, et demandent à être explicités par les individus lors des entretiens.

Les outils informatique sont, dans cette même perspective situés, entendu comme constitutifs du "réel" qu'ils donnent à voir au chercheur. Si l'on suppose généralement que le (jeune) chercheur travaille son corpus en fonction des théories qu'il a établies en amont et/ou des premiers résultats de terrain, et non en fonction des possibilités qui lui sont offertes par les logiciels, on souhaite ici insister sur le fait que, les limites des TIC étant progressivement repoussées, les logiciels offrent à présent une telle pluralité de fonctions que le chercheur enthousiaste est vite amené à vouloir tester toutes les possibilités de la plateforme.

Selon les fonctions qu'il peut utiliser dans tel programme, et qui ne seraient peut-être pas les mêmes dans un autre, la manière dont il va travailler son corpus via les fonctions disponibles l'amènera à constituer un corpus d'une manière qui aurait été différente s'il l'avait travaillé avec un autre logiciel. Dans l'étude qui nous occupe actuellement, si nous avons en amont pensé certaines catégories d'analyse, ce n'est que parce que *Sonal* offrait la possibilité d'allier analyse thématique et lexicométrique que l'on s'est proposé d'analyser non seulement le contenu mais également la forme discursive des verbalisations recueillies. C'est dans cette mesure que l'on se permet d'entendre l'outil comme élément primordial dans la constitution du corpus, et des résultats, tels qu'ils seront finalement présentés. Mais si l'outil prend autant d'importance à travers les possibilités techniques qu'il met à disposition du chercheur, ce dernier doit néanmoins rester la "tête pensante" de l'interaction entre le chercheur et le logiciel.

Le chercheur, 'tête pensante' de l'interaction homme-outil

Une fois importées dans *Sonal*, les verbalisations apparaissent comme un "matériau brut", dont on s'efforce de penser qu'il ne dit rien en soi, mais qu'il demande à être interprété par le chercheur, via une série de catégories d'analyse. L'outil de thématization par les codes couleurs du logiciel apparaît alors seulement comme une mise en forme, certes performante, mais qui ne remplace en rien la réflexion épistémologique qui sous-tend la compartimentation du contenu thématique du corpus, qui reste à l'entière charge du chercheur. L'établissement de catégories adéquates, à l'instar de tout le cadrage théorique du travail de thèse, s'effectue en amont et pendant la transcription et la re-lecture/ré-écoute des entretiens. Il apparaît évident qu'un logiciel ergonomique et simple d'utilisation favorisera la mise en place des catégories d'analyse, mais rappelons que cette opération aurait aussi bien pu se faire, de manière plus fastidieuse mais également "valable" en termes scientifiques, sur une feuille de papier, avec un crayon de bois. Nous avons ici pris l'exemple de la catégorisation des thématiques, mais il en va de même pour le choix des termes à indexer, et pour l'ensemble des analyses lexicométriques.

Pistes conclusives

C'est donc sur la capacité du chercheur à "penser", en amont et au-delà du logiciel qu'il utilise, que l'on souhaite conclure cette réflexion sur le rôle d'un outil comme *Sonal*. Cela n'empêche pas de constater les apports considérables de ce type de logiciels pour la recherche qualitative, en particulier en termes de confort de lecture, d'analyse et de diffusion. Mais l'enthousiasme du jeune chercheur, face à la diversité et l'attractivité grandissante des outils à sa portée, peut se trouver rapidement atténué lorsqu'il se trouve face à la tâche difficile de mettre en adéquation l'objectif précis de sa recherche avec les fonctions techniques du logiciel. C'est pourquoi, c'est finalement la question des limites, non plus des outils, mais des utilisateurs, qui reste à poser, en termes de temps, de moyens et de formation.

Références bibliographiques

Alber, A. 2010. « Voir le son : réflexions sur le traitement des entretiens enregistrés dans le logiciel Sonal ». *Socio-*

logos. *Revue de l'association française de sociologie* (5) (juillet 7). <http://socio-logos.revues.org/2482>

Albero, B. & F. Thibault. 2009. « La recherche française en sciences humaines et sociales sur les technologies en éducation ». *Revue française de pédagogie. Recherches en éducation* (169) : 53-66.

Bachelard, G. 2000. *La formation de l'esprit scientifique contribution à une psychanalyse de la connaissance*. Paris : J. Vrin.

Beacco, J-C. 2011. « Contextualiser les savoirs en didactique des langues et des cultures ». In *Guide pour la recherche en didactique des langues et des cultures*, 31-40. Paris : Archives contemporaines.

Bulea, E. & J-P Bronckart. 2012. « Les représentations de l'agir enseignant dans le cadre du genre "entretien" ». *Raido : Revista do PPG em Letras*, juin.

Clot, Y. & D. Faïta. 2000. « Genre et style en analyse du travail. Concept et méthodes ». *Travailler* (4) : 7-42.

Komis, V., C. Depover, & T. Karsenti. 2013. « L'usage des outils informatiques en analyse des données qualitatives ». <http://www.adjectif.net/spip/spip.php?article216>

Pothier, M. 2003. *Multimédias, dispositifs d'apprentissage et acquisition des langues*. Ophrys.

Rabardel, P. 1995. *Les hommes et les technologies : approche cognitive des instruments contemporains*. Armand Colin.

Yvon, F. & R. Garon. 2006. « Une forme d'analyse du travail pour développer et connaître le travail enseignant : l'autoconfrontation croisée ». *Recherches Qualificatives* 26 : 51-80.

[1] L'enthousiasme qui est ici manifesté pour le matériel vidéo en situation éducative ne fait pas oublier l'effort de distanciation nécessaire à l'analyse des données : les traces de l'activité ne doivent ainsi pas être entendues comme « *preuves irréfutables de la vérité de l'individu* » (Linard & Prax, 1984 : 186).

[2] L'autoconfrontation, ou *rétroaction vidéo* ou *rappel stimulé*, est entendue comme une méthode d'analyse de l'activité humaine consistant à confronter un ou plusieurs participants à leur propre activité en les incitant à commenter une vidéo de cette activité. La démarche d'autoconfrontation est principalement utilisée dans le cadre de l'ergonomie, de la psychologie du travail et des sciences de l'éducation. Pour une description de la méthode, voir Yvon & Garon (2006).

[3] Un très bon manuel en version PDF est téléchargeable avec le logiciel [ici](#)

[4] [des essais ont été faits sous Mac et Linux](#) mais aucune version stabilisée pour ces deux systèmes d'exploitation n'est disponible à ce jour