
L'élaboration et l'interrogation de bases de données orales comme enjeu pédagogique pour les Humanités digitales et les Sciences du langage. L'exemple de l'outil IMPACT

JACQUIN Jérôme *Université de Lausanne*

GRADOUX Xavier *Université de Lausanne*

Mot(s) clé(s): TIC - dispositif hybride - bases de données - oralité - analyse conversationnelle -

1. Introduction

La présente contribution entend faire le point sur le projet IMPACT, un outil présenté une première fois à l'occasion de l'édition 2012 du colloque de l'AIPU (Gradoux & Jacquin, 2012, 2013)[1]. Rappelons qu'IMPACT – pour *Interface Multimedia : Présentation – Analyse – Commentaire* – est un dispositif techno-pédagogique hybride, c'est-à-dire mêlant apprentissage en présence et à distance (Charlier, Deschryver, & Peraya, 2006) et visant l'amélioration et l'enrichissement du scénario pédagogique initial par l'usage de TIC (Karsenti & Larose, 2001). Plus concrètement, IMPACT est une interface multimédia accessible et éditable en ligne (Alessi & Trollip, 2001) et permettant de construire des fiches rassemblant dans un même espace de consultation et d'édition différentes ressources hétérogènes, telles qu'une source audiovisuelle, sa transcription et son analyse.

Dans cette contribution, on tâchera plus précisément de rendre compte de l'évolution pédagogique du projet, en mettant un accent particulier sur un usage émergent de l'outil et sur ses conséquences tant pour l'enseignement que pour la recherche. L'utilisation croissante d'IMPACT dans notre domaine de spécialisation – les Sciences du langage – nous a en effet conduits à réfléchir, plus généralement, sur la manière d'intégrer enseignement et recherche dans l'élaboration et l'interrogation de bases de données orales numérisées. En ce qu'elle prend acte des rapports réflexifs et mutuellement déterminants qui lient sources, contenus et technologies, notre contribution s'inscrit ainsi dans le courant de réflexion des *Humanités digitales*.

Après un bref rappel de l'objectif initial de l'outil et de ses fonctionnalités principales (2.), nous présentons l'usage émergent d'IMPACT en tant qu'il permet et favorise la constitution de bases de données orales (3.). Nous opérons ensuite une comparaison systématique avec plusieurs bases de données orales francophones disponibles (4.) et concluons cette contribution avec quelques perspectives.

2. Objectif initial et fonctionnalités principales

L'enseignement qui a motivé le développement d'IMPACT s'inscrit dans les Sciences du langage et plus précisément dans l'analyse conversationnelle[2]. En tant que discipline spécifique, l'analyse conversationnelle met un accent particulier sur le caractère mutuellement constitutif de la confection des données (qui relèvent d'enregistrements transcrits de conversations ordinaires), de leur analyse (qui se caractérise par une grande attention portée aux détails

de ces conversations) et de la production du discours théorique (qui opère par la description de régularités telles qu'elles sont appréhendables par des regroupements systématiques). L'analyse conversationnelle se veut donc éminemment descriptive et inductive, privilégiant un contact direct et constant aux données.

L'obstacle majeur auquel l'enseignement était confronté résidait dans la manipulation constante et fastidieuse de fichiers à la fois hétérogènes et dispersés – sources audio-visuelles, transcriptions d'extraits, analyses de ces extraits, documents théoriques – nécessitant une consultation et une édition dans des logiciels différents et dans des fenêtres distinctes, avec les constants problèmes de superposition que cela implique. Ces problèmes ergonomiques s'accompagnaient en outre d'une dispersion du contenu, conduisant les étudiants comme les enseignants à mettre de côté la source originale au profit de sa seule transcription. Finalement, la situation initiale provoquait rapidement une multiplication de versions pour chaque fichier, ce qui nuisait fortement à la collaboration entre enseignants et entre étudiants et enseignants.

Citant notre contribution de 2012,

« Il manquait un outil permettant de rendre la consultation des documents ergonomique et d'éviter que la dimension technique ne vienne nuire inutilement à l'apprentissage. Il fallait un outil capable (i) d'afficher les documents nécessaires dans un même espace visuel, (ii) d'ouvrir chacun de ces documents sans qu'aucune installation de logiciel spécifique ne soit nécessaire, (iii) de répercuter automatiquement toute modification sur les consultations ultérieures, (iv) d'assurer la pérennité et l'accessibilité à distance de la matière. » (Gradoux & Jacquin, 2012).

La solution a donc été de développer une application web, IMPACT, permettant de générer des fiches au sein desquelles les éléments pertinents (une source audio/vidéo, sa transcription, des annotations théoriques ou analytiques) sont organisés de manière systématique et ergonomique, accessibles et éditables à distance via un simple navigateur et diffusables par vidéo-projection pour les séances en coprésence :

21. Marques et stratégies d'introduction

type: fiche de cours état: archivée date de présentation: 22/03/2012

Source



Théorie

L'introduction d'un nouveau topic est souvent rendue visible par un ensemble de marques linguistiques qui rendent reconnaissable le topic en tant que tel. Ce marquage sera d'autant plus fort, si le nouveau topic est en discontinuité avec le topic précédent ou s'il est peu prédictible.

Il s'agit ici de marques produites par un interactant pour rendre saillant l'élément dont il parle. Mais ce topic marqué n'accèdera pas forcément au rang de topic interactionnel: il peut ne pas être ratifié par les autres interactants.

Il y a tout d'abord des marqueurs préliminaires qui servent à embrayer l'acte de discours: ils sont de type phatique (dis donc), cognitif (tu sais) ou métalinguistique (je voulais te dire).

Ensuite, pour introduire véritablement un topic, on relève des marqueurs topicaux qui servent à ancrer le thème dans la discussion: il s'agit de faire mention

Transcription

Contexte : après 2h du débat de l'entre-deux-tours entre Nicolas Sarkozy (NS) et Ségolène Royal (SR), diffusé sur TF1 et France 2 en direct le 2 mai 2007.
Journalistes : Arlette Chabot (AC) et Patrick Poivre d'Arvor (PPD)

```

1 NS et la certitude de la réalisation de cette promesse (.) ça s'ra
2 le droit opposable (.) et la capacité]
3 SR [mais] (.) est-ce-
4 NS =d'aller devant un tribunal pour faire valoir ses [droits]
5 SR [qu'on] peut-
6 NS =vous voyez madame/ (.) c'est ni ridicule\ (...) ni accessoire\
7 (...) c'est peut-être même ce qui fait la différence/ (...) entre
8 la vieille politique/ (...) et la politique mo[derne\
9 [alors vous avez
10 réussi tous les deux [une prouesse]
11 SR [attendez] (.) j'ai quelque chose à
12 [dire (.) parce que là je pense (.) je pense]
13 PPD [justement oui (.) le temps d'antenne]
14 AC [un un mot puis euh et on enchaîne]
15 SR &que l'on atteint le summum de l'immoralité politique\ (...) (h)
16 je suis scandalisée de ce que je viens d'entendre (...) parce que
17 jouer (...) avec le handicap (...) comme vous venez de le faire
18 (...) est proprement scandaleux\ (...) pourquoi\ (...) lorsque
19 j'étais ministre de l'enseignement scolaire\ (...) c'est MOI qui ai
20 créé le plan HANDISCOL

```

Exemplification

En 9, les modérateurs tentent de prendre la parole en anticipant un peu le PTP du «discours» de NS. En 11 SR, qui avait par ailleurs produit des «claim for the turn units» en 3 et 5, produit en chevauchant PPDA: attendez (.) j'ai quelque chose à dire (.) parce que là je pense (.) je pense. Cet

Fig. 1. Capture d'une fiche Impact

- Ainsi, parmi les principales fonctionnalités d'IMPACT, il est possible de :
- Diffuser une source audio ou audio-vidéo dans son intégralité ou en choisissant un extrait.
- Piloter la lecture de la source au clavier (lecture/pause ; ralenti ; retour de 2 sec. ; rembobinage rapide).
- Transcrire la source ou un extrait en respectant les standards du domaine, mais dans un canevas intuitif et ergonomique.
- Exporter la transcription ou en importer une existante.
- Synchroniser ou non le défilement de la transcription et de la source.
- Ajouter des commentaires textuels (théoriques ou analytiques).
- Joindre des documents PDF ou des images.
- Créer un scénario pédagogique grâce à une logique d'étapes, permettant à l'enseignant d'évaluer progressivement le travail de l'étudiant.
- Coordonner chaque étape à un rendu de devoirs dans la plateforme d'apprentissage Moodle.

3. Usage émergent d'IMPACT en tant qu'outil de base de données

Ce dispositif, au départ principalement motivé par des besoins ergonomiques et techniques, a eu un certain nombre de conséquences, notamment pédagogiques. En 2012, nous en identifions trois : (i) l'enrichissement du scénario pédagogique, au travers de la réalisation, par les enseignants comme par les étudiants, de fiches multimodales intégrant théories et pratiques dans un même espace de consultation et d'édition (ii) l'autonomisation, saluée par les

étudiants, de l'enseignement vis-à-vis de la situation de coprésence et (iii) l'influence positive sur la motivation et l'engagement des étudiants.

Deux ans plus tard, une nouvelle conséquence est clairement apparue, au fur et à mesure d'une utilisation croissante d'IMPACT. Celle-ci a trait à l'utilisation émergente du dispositif en qu'outil de constitution et d'interrogation de bases de données.

Progressivement, au fil des semestres et des travaux personnels effectués par les étudiants, le stock d'extraits s'est en effet largement enrichi. Les étudiants participent ainsi constamment à documenter des pratiques à la fois authentiques et locales – éphémères par définition. Cette réserve d'extraits est également devenue une ressource au sein de laquelle les enseignants et les chercheurs sélectionnent des exemples, voire opèrent des analyses plus systématiques. Cet usage d'IMPACT en tant que ressource documentaire, n'a pas été pensé, réfléchi, prévu ou encore outillé à l'avance. Il s'agit d'un bricolage émergent, à toutes fins pratiques. Il en résulte que l'outil souffre de certaines limites (cf. infra 4.), mais là réside aussi son intérêt dans le sens où cette plasticité permet à l'outil de s'adapter aux besoins, aux contextes, au gré du travail de concertation mené avec les étudiants. La conclusion sera l'occasion de développer quelques perspectives et améliorations relatives à ce point.

Plus concrètement, les trois exemples suivants témoignent de la diversité de l'utilisation actuelle d'IMPACT en tant qu'outil de base de données :

1. L'enseignement à l'origine d'IMPACT – « introduction à l'analyse conversationnelle » – compte désormais un peu plus de 120 extraits sonores de 2 à 3 minutes. Cela peut paraître peu, mais ces extraits originaux ont été enregistrés et transcrits de manière standardisée par les étudiants, puis vérifiés par les enseignants, ce qui en fait des éléments précieux pour la recherche sur les interactions quotidiennes.
2. Dans le cadre d'un enseignement consacré à l'acquisition du langage, plus précisément de l'acquisition de la compétence narrative, un de nos collègues a enregistré des cours d'anglais au centre de langues de l'Université de Lausanne. Le dispositif pédagogique en question invite chaque étudiant du cours de langue à raconter une même histoire quatre fois, mais dans un temps de plus en plus court. Huit étudiants étaient concernés, pour un total de 32 extraits. Notre collègue a mis à disposition les extraits sur IMPACT et les étudiants se sont chargés de la transcrire et de les analyser dans le cadre de leur travail personnel. IMPACT a ainsi permis de documenter facilement une expérience complète d'enseignement et de la rendre disponible pour différentes analyses, linguistiques ou non.
3. Finalement, pour clore ce panorama des usages émergents d'IMPACT, on peut noter l'utilisation de l'outil pour une recherche doctorale consacrée aux consultations médicales, analysées du point de vue linguistique. Le doctorant a enregistré 35 consultations de médecine générale en suisse romande pour une quinzaine d'heures d'enregistrement. IMPACT accompagne le doctorant durant toute sa recherche, pour non seulement pour stocker, transcrire ou organiser ses données, mais aussi pour présenter sa recherche et enseigner sur la base de son corpus.

4. Comparaison avec des bases de données existantes

Prenant acte de cette situation, nous opérons maintenant une comparaison des possibilités offertes par l'outil avec certaines des bases de données actuellement exploitées en sciences du langage francophones. Cette comparaison vise, d'une part, à évaluer le bien-fondé de cette utilisation détournée d'IMPACT et à identifier des solutions alternatives potentiellement plus efficaces. Il s'agit, d'autre part, d'orienter les développements futurs d'IMPACT en prenant exemple sur des outils spécialisés.

Trois bases de données ont été considérées : (i) CLAPI[3] (Corpus de langue parlée en interaction) est une base importante créée et hébergée par le laboratoire ICAR à l'Université Lyon 2, qui se spécialise dans l'analyse des interactions authentiques quotidiennes ou institutionnelles ; (ii) ESLO[4] (Enquêtes SocioLinguistiques à Orléans) regroupe des enregistrements de productions verbales diverses (entretiens, conférences, interactions dans la rue, etc.), qui constituent « un formidable témoignage sur la ville et sur le français et les langues parlées quotidiennement dans toutes leurs variétés et leur diversité »[5] ; (iii) VALIBEL[6] (Variétés Linguistiques du français en Belgique) « vise [...] à documenter la variation linguistique en Belgique Francophone » (Dister, Francard, Hambye, & Simon, 2009, p. 115).

Cinq critères de comparaison ont été retenus : la taille, l'accessibilité, l'ajout d'annotations et de méta-données, la recherche et la consultation.

Avec une trentaine d'heures de données enregistrées, IMPACT est d'une taille relativement modeste en comparaison avec les autres bases puisque ESLO et VALIBEL revendiquent chacune plus de trois cents heures. IMPACT est toutefois alimenté par les étudiants, ce qui garantit une croissance de la base de semestre en semestre et valorise les travaux des étudiants en tant qu'ils contribuent à un projet d'enseignement et de recherche plus large.

L'accès aux données varie sensiblement d'une base à l'autre, oscillant entre des modèles complètement ouverts et d'autres bien plus restrictifs. IMPACT est le plus fermé de tous dans le sens où les utilisateurs n'ont accès qu'aux fiches des cours dans lesquels ils sont inscrits.

Les trois bases de données fournissent en outre des méta-données – en général fondées sur des standards tels que le TEI (*Text Encoding Initiative*[7]) – sur l'enregistrement, les locuteurs, le type de transcription, etc. ainsi que des annotations locales comme la notation de chevauchements. Le cas d'IMPACT est à part : la possibilité d'ajouter des annotations dans les champs textuels existe, mais ces annotations ne sont ni obligatoires, ni standardisées. C'est à l'étudiant ou à l'enseignant/chercheur de décider du type et de la forme de l'information qu'il désire fournir.

Tant CLAPI, ESLO que VALIBEL offrent la possibilité d'interroger la base, soit sous forme d'un tri par méta-données, soit par une recherche, dans la transcription, de chaînes de caractères. IMPACT ne dispose pas de telles fonctions de recherche, mais permet aux utilisateurs de créer un parcours ordonné de notions. Cette possibilité de classement, couplé aux annotations, est appréciable pour des usages pédagogiques, mais est moins adaptée à la recherche.

Finalement, IMPACT surclasse les autres projets pour ce qui est de la présentation et de l'édition des données,

puisqu'il permet non seulement de visualiser simultanément la source, la transcription et les annotations, mais aussi de transcrire et d'annoter directement dans l'outil.

En somme, IMPACT est limité en tant qu'outil de base de données : il ne permet pas de faire des recherches et ne dispose ni de méta-données standardisées ni d'un accès ouvert. Il offre en revanche des possibilités d'adaptation à différents contextes, une consultation agréable et une édition intégrée.

5. Perspectives

Dans cette contribution, nous avons commencé par rapidement rappeler les besoins et objectifs à l'origine du développement et des fonctionnalités d'IMPACT. Nous avons insisté ensuite sur un usage émergent de l'outil, celui d'organiser les données sous forme de bases documentaires. Cet usage a motivé une comparaison systématique d'IMPACT avec trois projets francophones bien connus. La comparaison a fait clairement apparaître les limites actuelles d'IMPACT, mais aussi un avantage, rendu possible par la jeunesse et la plasticité du projet : IMPACT permet et favorise la constitution progressive et collaborative, par les enseignants comme par les étudiants, de bases de données orales enrichies et disponibles aussi bien pour l'enseignement que pour la recherche.

Toutefois, pour pallier les limites identifiées, de nouveaux développements sont planifiés :

- La possibilité d'interroger la base de données en opérant des recherches textuelles au sein des champs transcription.
- L'autonomisation vis-à-vis de Moodle, pour un accès plus ouvert et des usages qui dépassent l'enseignement.
- Une standardisation des méta-données, sous forme de mots-clés introduits collaborativement par les étudiants et les enseignants.

Ces modifications permettraient à terme d'améliorer l'enseignement en facilitant la découverte et la sélection d'extraits représentatifs, de valoriser davantage la participation des étudiants à l'enseignement et à la recherche et, dans le même mouvement, d'encourager le partage des données au sein de la communauté.

Références

Alessi, S. M., & Trollip, S. R. (2001). *Multimedia for learning: methods and development*. Boston: Allyn & Bacon.

Baude, O., & Dugua, C. (2011). (Re)faire le corpus d'Orléans quarante ans après : quoi de neuf, linguiste ? *Corpus*, 10, 99–118.

Bert, M., Bruxelles, S., Etienne, C., Mondada, L., & Traverso, V. (2009). Exploitation de la plateforme Corpus de Langue Parlée en Interaction (CLAPI) : le cas de "voilà" dans les chevauchements. *Cahiers de Linguistique*, 33(2), 243–268.

Charlier, B., Deschryver, N., & Peraya, D. (2006). Apprendre en présence et à distance : Une définition des dispositifs hybrides. *Distances et Savoirs*, 4, 469–496.

Dister, A., Francard, M., Hambye, P., & Simon, C. (2009). Du son, du texte, des métadonnées: l'évolution de la banque de données textuelles orales VALIBEL (1989 - 2006). *Cahiers de Linguistique*, 33(2), 113–129.

Gradoux, X., & Jacquin, J. (2012). La résolution d'un problème d'enseignement par l'usage de TIC ainsi que sa plus-value pédagogique. In *Actes du 27e congrès de l'Association internationale de pédagogie universitaire (AIPU)*.

Gradoux, X., & Jacquin, J. (2013). Le projet IMPACT. Retrieved from www.unil.ch/impact

Hutchby, I., & Wooffitt, R. (2008). *Conversation Analysis*. Cambridge: Polity Press.

Karsenti, T., & Larose, F. (Eds.). (2001). *Les TIC... au coeur des pédagogies universitaires*. Québec: Presses de l'Université du Québec.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), 696–735.

Ten Have, P. (2007). *Doing Conversation Analysis*. London: Sage.

[1] IMPACT est en développement depuis 2010. En 2011, il a bénéficié du soutien du *Fonds d'Innovation Pédagogique* de l'Université de Lausanne (www.unil.ch/fip). Depuis 2012, il est financé conjointement par le *Réseau Interfacultaire de Soutien Enseignement et Technologie* (www.unil.ch/riset) et par la Faculté des lettres (www.unil.ch/lettres).

La présente publication bénéficie pour partie du soutien du Fonds National Suisse de la recherche (FNS), par le biais d'une bourse *Early Postdoc.Mobility* octroyée à Jérôme Jacquin pour un séjour de recherche de 12 mois à la *Victoria University of Wellington*, Nouvelle Zélande (N° P2LAP1_155110).

[2] Voir l'article fondateur de Sacks, Schegloff & Jefferson (1974). Pour une initiation à la discipline, voir plutôt Ten Have (2007) et Hutchby & Wooffitt (2008).

[3] <http://clapi.univ-lyon2.fr/>, consulté le 14 août 2014 (voir aussi Bert, Bruxelles, Etienne, Mondada, & Traverso, 2009).

[4] <http://eslo.huma-num.fr/>, consulté le 14 août 2014 (voir aussi Baude & Dugua, 2011).

[5] Site web ESLO, consulté le 4 août 2014.

[6] <http://www.uclouvain.be/valibel>, consulté le 14 août 2014 (voir aussi Dister et al., 2009).

[7] <http://www.tei-c.org/index.xml>, consulté le 14 août 2014.