

Rapport final du projet Mulce

Version du 17 mars 2011

A. Identification

Programme – année	ANR-06-CORP-006
Projet (acronyme)	Mulce
Titre complet du projet	Échange de corpus d'apprentissage multimodaux
Coordinateur du projet Nom, laboratoire, organisme de rattachement	Thierry Chanier
Partenaire 1 Nom du responsable, laboratoire, organisme de rattachement	Thierry Chanier Laboratoire de Recherche sur le Langage (LRL), Université Blaise Pascal, (Laseldi et UFC en début de projet)
Partenaire 2 Nom du responsable, laboratoire, organisme de rattachement	Laboratoire d'Informatique de l'Université de Franche-Comté, Christophe Reffay, (en détachement à l'INRP->ENS-Lyon, UMR STEF, ENS-Cachan)
Période du projet (date début – date fin)	22/01/2007 - 21/12/2010
Rapport confidentiel (OUI/NON)	Non
Date de fin de confidentialité	

Rédacteur de ce rapport

Civilité, prénom, nom	M. Thierry Chanier
Téléphone	04 73 34 68 39
Adresse électronique	thierry.chanier@univ-bpclermont.fr
Date de rédaction	01/03/2011
Civilité, prénom, nom	M. Christophe Reffay
Téléphone	01 47 40 76 15
Adresse électronique	Christophe.Reffay@ens-cachan.fr
Date de rédaction	01/02/2011

B. Rapport scientifique

B.1 - Résumé du travail accompli

La recherche sur les interactions en ligne en situation d'apprentissage offre encore trop peu souvent la possibilité d'accéder aux données à partir desquelles les chercheurs ont élaboré les analyses présentées dans les publications. Cela restreint, d'une part, la compréhension des phénomènes étudiés et, d'autre part, empêche toute répllication dans le but de comparaisons, d'analyses cumulatives ou contrastives. Dans le projet Mulce, nous défendons le point de vue méthodologique suivant : pour permettre une analyse des interactions situées, il convient de relier les différentes données issues de formations en ligne pour construire un objet d'analyse, exploitable par différentes équipes et disciplines. Le constat actuel est que les données sont souvent décontextualisées, parcellaires ou simplement inaccessibles à la communauté des chercheurs.

Nous proposons donc de structurer les données en corpus d'apprentissage (*LEarning & TEaching Corpora*, LETEC) de façon à rendre possible leur échange et la capitalisation des analyses. Le protocole de recherche, le scénario pédagogique, les interactions, productions et traces, les licences et les analyses capitalisables en sont les constituants. Une structure de données a été définie (Mulce-struct, 2010). Elle assemble dans une enveloppe au standard IMS-CP, les éléments précédents décrit en XML avec les ressources associées. Une structure de métadonnées a également été définie pour chaque corpus suivant les recommandations OLAC (*Open Language Archives Community*). La question de l'éthique a fait l'objet d'une attention particulière. Des procédures d'anonymisation ont été appliquées sur les corpus et les licences intégrées à la structure de corpus.

Une banque de corpus a été constituée (Mulce-databank, 2011). Elle permet l'accès libre aux données suivant le protocole OAI (*Open Archive Initiative*). 6 corpus d'apprentissage y ont été déposés, correspondant à 6 formations en ligne. A cela viennent s'ajouter 26 corpus distinguables (Mulce-contenu, 2011). Un tel corpus est tout à la fois un sous-corpus du corpus d'apprentissage et un corpus en soi. Il en existe de 3 types : corpus associant une publication scientifique et les données associées ; rassemblant des données prêtes à l'analyse avec la mise en forme pour des outils/logiciels libres ; partageant des analyses avec des outils associés. Cette banque est aujourd'hui déclarée au sein de l'archive OLAC (Olac-Mulce, 2011). Nous y avons défini un nouveau type de corpus (Olac-Letec, 2010). Un site de documentation développe toute notre démarche méthodologique (Mulce.org, 2011).

Des contacts réguliers se sont établis avec des équipes de chercheurs extérieurs permettant d'utiliser leurs outils d'analyse des interactions avec nos données structurées. Enfin nous avons participé ou organisé des symposiums ou ateliers dans des conférences internationales rassemblant les chercheurs travaillant sur les données ou outils d'analyse des interactions en ligne en situation d'apprentissage (Eurocall-Mulce, 2010).

Abstract

Research into online learning, whether aimed at understanding this form of situated human learning, or at evaluating relevant pedagogical scenarios and settings, or at improving technological environments, requires interaction data, from all participants in the learning situations, to be made available. At present, interaction data are often either inaccessible, or only partly accessible to researchers not involved in the original projects. Moreover, data are fragmented and therefore decontextualized with respect to the original teaching / learning settings and may be buried in a proprietary format within the technological environment. Consequently, research lacks a scientific basis. In literature, comparisons are often attempted between objects that are ill-defined and which may in fact differ. The processes of scientific enquiry, such as re-analyzing, replicating, verifying, refuting or extending the original findings, are therefore impossible.

To address this inconsistency, we have created and disseminated a new type of corpus. This is a contextualized learner corpus, entitled "*LEarning and TEaching Corpus*" (LETEC) (Olac-Letec, 2010). Such corpora include not only the data that correspond to the output of learner activity during online courses, but also the data that correspond to the contexts for such output. Sharing LETEC corpora within the research community implies the following three prerequisites: (1) corpora are formatted and structured according to a new model which is compatible with existing standards for corpora and for learning design specifications (Mulce-struct, 2010); (2) corpora are placed on a server offering cross-platform compatibility (according to the *Open Language Archives Community* – (OLAC-Mulce(2010)) and free access; (3) an ethical protocol, as well as copyright licenses, are formulated.

Mulce databank (2011) includes 6 LETEC corpora and 26 distinguishable corpora (Mulce-contenu, 2011) of 3 types: corpus designed as replication data sets (Gary, 2007) which associate a publication and its specifically formatted data, i.e. corpus including data structured in a way they can be directly used with open access analysis tools; corpus with analysis included and associated tools. We developed the information site (Mulce.org, 2011), which presents our methodological approach.

Regular meetings with external research teams offered occasions to start benchmarking analysis tools and our data. Eventually we attended or have organized international workshops which gathered researchers working on data and analysis tools of online interactions occurring in learning situations (Eurocall-Mulce, 2010).

B.2 - Rapport scientifique proprement dit

Avertissement : les index de référence à nos travaux sont de deux natures : un index entre crochets (ex.[com1]) renvoie aux références de publications de la section C.1 ; un index entre parenthèse avec date (ex. (Olac-Mulce, 2011)) renvoie aux références Mulce de la section C.4

1. Enjeux, situation du sujet, objectifs et problématique

Le bilan qui a été fait lors du symposium en janvier 2005 à Amiens dans [HC 05]¹ montre que les travaux des domaines se rapportant à l'utilisation des technologies dans l'apprentissage (EIAH, TICE, CSCL, AL&SIC²) sont assez peu reconnus scientifiquement pour différentes raisons : les chercheurs en psychologie ou en sciences de l'éducation reprochent bien souvent le manque de méthodologie dans le protocole d'expérimentation, d'autre le manque de répliquabilité ou de validité des résultats avancés. Il est vrai que les expérimentations de formation en ligne se multiplient, facilitées par la concrétisation récente de nombreux dispositifs, en particulier dans les universités ouvertes. De nombreuses universités classiques leur emboîtent le pas et s'ouvrent à la formation en ligne, le plus souvent en accompagnement du présentiel, mais parfois même pour des diplômes entièrement à distance. Ainsi, chaque équipe de recherche trouve dans son environnement un cadre lui permettant de construire ses propres expérimentations.

Le protocole de recueil des données n'est pas toujours très formel, les données sont souvent partielles, mais surtout, quand les résultats sont rapportés dans les communications scientifiques, ils sont épurés de leur contexte et ne souffrent donc aucune critique sur leur interprétation. Enfin, le contexte global du dispositif de formation n'est jamais disponible pour le lecteur étranger à l'expérimentation. Dans ces conditions, il est bien difficile de comparer les méthodes ou d'ouvrir le débat scientifique sur les résultats d'analyse. Et puisque les outils d'analyse sont dépendants de chaque contexte (voire de chaque expérimentation), on comprend qu'ils aient des difficultés à dépasser le statut de prototype.

La situation est identique dans le sous-domaine s'intéressant aux environnements multimodaux d'échanges synchrones en ligne. Ils suscitent l'engouement de plus en plus de formateurs et d'apprenants depuis une décennie. Ces plateformes de formation générant des interactions complexes entre les participants renouvellent la recherche sur les interactions en ligne en situation d'apprentissage, notamment le questionnement sur les traces et leur traitement. En outre, la multimodalité, pourvue de nouveaux atouts, a suscité de nombreuses analyses qualitatives et quantitatives, motivées par les types originaux de dialogues qu'elle favorise, par le soutien qu'elle offre dans la communication en langue-cible.

¹ Henri F., Charlier, B. "L'analyse des forums de discussion pour sortir de l'impasse". In G.-L. Baron, E. Bruillard, and M. Sidir (Dir.), editors, *Symposium, formation et nouveaux instruments de communication*, Amiens, France, January 2005

² CSCL, Computer-supported Collaborative Learning ; AL&SIC, pour Apprentissage des Langues et Systèmes d'Information et de Communication, EIAH ; Environnements Informatiques pour l'Apprentissage Humain ; TICE : Technologies d'Information et de Communication pour l'Education).

Cependant, malgré leur nombre, les publications rapportant ces analyses ne permettent pas d'établir de comparaison entre ces environnements. De plus, la méthodologie de recherche habituellement utilisée dans l'analyse des interactions en ligne, qu'elles soient synchrones ou asynchrones, ne prend souvent en compte qu'une fraction des données, voire qu'un seul type d'interaction privilégiant un outil spécifique (quelques forums, quelques extraits de séances synchrones), et ne les organise pas de façon systématique. L'absence de pratiques partagées et de guides méthodologiques concernant l'élaboration de corpus recueillant ces interactions, tout comme l'absence de corpus organisés à des fins d'évaluation des différentes approches freinent la recherche. Au premier abord, des raisons évidentes semblent expliquer ces manques. Les corpus peuvent être perçus comme coûteux à collecter et difficiles à organiser. En effet, de multiples niveaux d'annotations s'offrent aux chercheurs du fait que modes et modalités interagissent constamment pour convoquer des actions à caractères verbales ou non verbales ; ces actions peuvent elles-mêmes être organisées dans des schèmes à caractères communicatifs, ou collaboratifs.

Certaines communautés de chercheurs ont cependant décidé de relever le défi, comme celle centrée sur l'apprentissage collaboratif en ligne (CSCL). Il s'y développe un intérêt grandissant pour la méthodologie de structuration de corpus et la définition de formalismes pivots, motivées par les perspectives d'échanges, de partage des données, des analyses et des outils. En ce qui concerne la communauté étudiant l'apprentissage des langues en ligne (AL&SIC), l'intérêt qu'elle accorde à la communication multimodale ou aux interactions en ligne ne l'a pas encore attirée vers ces questions méthodologiques. La comparaison avec d'autres domaines plus traditionnels des sciences du langage montre cependant que leur prise en considération va de pair avec une recherche mieux structurée.

En ce qui concerne l'association publication et données / analyses ayant conduit à ces publications, certaines communautés en sciences humaines ou sociales ont ouvert la voie en développant la notion d'ensemble de données pour la réplcation :

Replication data sets include the original data and any other information needed to reproduce the numerical results in a published work. [...] making publicly available a replication data set for each of their empirical articles or books. Citation credit should be apportioned both for the original article and separately for the data. (Gary, 2007 : 145)

Elles ont également développé les infrastructures matérielles et logicielles pour associer publication et partage de données de recherche, tel le réseau DataVerse³.

L'objectif du projet Mulce est de contribuer à améliorer la scientificité des recherches sur les interactions en ligne en situation d'apprentissage dans les domaines EIAH, CSCL et AL&SIC en

- permettant la réutilisation de données de recherche (pérennité, partage) en protégeant les acteurs de la situation d'apprentissage ;
- proposant un cadre pour organiser et décrire un corpus de données issues d'une situation d'apprentissage (collaboratif) en ligne ;

³ Dataverse (2009). *Site de la communauté Dataverse Network développant un réseaux de banques de données de recherche associées aux publications*. [site Internet]. Harvard University. <http://thedata.org/>

- concevant et mettant en œuvre une plateforme pour rendre ces corpus accessibles, interrogeables et téléchargeables ;
- assurant une visibilité internationale à ces corpus ;
- développant et intégrant des analyses réalisées sur ces données à partir d'outils mis au point par les chercheurs de ces domaines.

2. Matériels (terrains, sources, données...), méthodes et travaux réalisés

Certaines expérimentations de formation en ligne avaient eu lieu avant que ne commence le projet Mulce. Durant le projet Mulce, ont eu lieu l'organisation de leurs données, leurs transformations, les transcriptions de vidéos.

Simuligne et Copéas sont les deux situations d'apprentissage en ligne à partir desquels nous avons travaillé pour constituer deux corpus d'apprentissage et plus de 20 corpus distinguables.

Les données primaires avant Mulce étaient les suivantes :

- Expérimentation Simuligne : 30 000 fichiers répartis dans 2 708 dossiers correspondant, entre autres à : Interactions: 2 686 mess. forum, 4 062 courriels, 5 680 tours de clavardage ; Productions : 93 doc. textuels, 28 fichiers audio ; Productions affichées : 342 pages web incluant 115 images et 44 fichiers audio ; guide apprenant, guide tuteur, guide natifs.
- Expérimentation Copéas : 37 vidéos (27h), 512 autres fichiers dans 117 dossiers, 180 000 lignes de traces et transcription dans Mulce-struct. Ceux-ci correspondent à : Interactions : 5 506 tours de parole audio (8h29 en temps cumulé), 1 529 tours de clavardage, 16 séances Lyceum ; Protocole recherche : 14 questionnaires apprenants, 9 entretiens, 9 entretiens d'auto confrontation ou "*Critical Event Recall*" (vidéos et audio) ; guide apprenant.

La section 7. *Modalités de création* présente plus précisément la méthodologie et les formalismes que nous avons développés pour transformer ces données en corpus. Ceci nous a amené en particulier à stabiliser nos conventions de transcriptions multimodales (Mulce-transcript , 2009), développer la structure XML Mulce-struct (2010), produire des procédures d'anonymisation, étendre l'utilisation du formalisme de scénario IMS-LD à la définition des protocoles de recherche, utiliser les métadonnées OLAC et Dublin Core et étendre la typologie OLAC des corpus (OLAC-Letec, 2010), développer des méthodes de conversions automatiques des blogues dans les structures Mulce, des structures d'interactions Mulce vers les structures Calico (forum) ou Tatiana (transcriptions et vidéos pour alignement et analyses).

En étudiant la question de l'éthique et des droits, nous avons rédigé une licence d'utilisation (Mulce-licence, 2010), stabilisé la formulation d'un contrat de consentement éclairé et explicité les procédures de respect de l'éthique (Mulce-ethique, 2010).

Nous avons ouvert un lieu de dépôt et la collection correspondante (*repository*) connectés à OLAC (*Open Language Archive Community*) (OLAC-Mulce, 2011) et développé la banque de corpus (Mulce-databank, 2011) en accès libre.

3. Résultats scientifiques obtenus, présentation de leur caractère original, mise en perspective avec les attentes initiales et l'état de l'art

Nous avons défini pour la première fois dans le milieu de la recherche sur les situations d'apprentissage en ligne, particulièrement en environnements multimodaux, une façon de systématiser le recueil des données, de les structurer pour les transformer en un véritable objet d'étude scientifique qui peut être partagé et faire l'objet d'analyses croisées, contradictoires ou encore évaluer la qualité des outils d'analyse disponibles (*benchmarking*).

Cette structure se définit ainsi : Un corpus d'apprentissage (LETEC) assemble de façon systématique et structurée un ensemble de données, particulièrement d'interactions, et de traces issues d'une expérimentation de formation partiellement ou totalement en ligne, enrichies par des informations techniques, humaines, pédagogiques et scientifiques permettant leur analyse en contexte.

- Le dispositif pédagogique peut-être librement décrit, mais il est préférable de le faire de façon détaillée en précisant le **scénario pédagogique**, les différents rôles des participants, en particulier des apprenants et enseignants, ainsi que les environnements technologiques retenus avec leurs fonctionnalités, et leurs caractéristiques dédiées aux interactions.
- De la même façon, si l'expérimentation inclut un **protocole de recherche**, le rôle des chercheurs, le séquençement des activités afférentes (administration de questionnaires, entretiens, etc.) seront utilement décrits.
- Les deux parties précédentes correspondent à ce qui était prévu avant le déroulement de la formation, c'est-à-dire à un modèle. Suivant la terminologie des langages objets, le modèle "s'instancie" lors de l'acte pédagogique (avec tous les changements inopinés afférents). La partie **instanciation** assemble donc, d'une part, les enregistrements des interactions des participants (sous forme textuelle, audio ou vidéo), leur productions individuelles (tels que les travaux écrits ou oraux, les journaux de bord) et, le cas échéant, les traces système (temps de connexion, statistiques de participation, etc.). Elle regroupe d'autre part, le cas échéant, les questionnaires remplis, les enregistrements d'entretiens et matériaux afférents (grille d'entretien, matériaux pour auto-confrontation, etc.)
- La partie publique de la **licence** donne accès aux licences d'utilisation du corpus par la communauté de chercheurs et de praticiens et les formulaires de contrat d'éthique remis aux participants. La partie privée de la licence n'est pas directement intégrée au corpus, mais conservée par le responsable du corpus. Elle incorpore notamment les patronymes et coordonnées des participants, ainsi que les contrats (consentements éclairés) signés.
- Les **analyses** ne font pas en général partie du corpus d'apprentissage, mais seront adjointes ultérieurement sous la forme de corpus distinguables (voir la section suivante). Quant au cas intermédiaire des transcriptions des enregistrements vidéo ou audio, nous les avons intégrées dans les corpus déposés dans la plateforme Mulce, en sachant qu'elles peuvent être recommencées ou modifiées. La figure 1 fait apparaître la partie analyse comme l'objectif orientant l'ensemble de l'effort de collecte et d'organisation.

Pour connaître l'organisation d'un corpus Mulce et le détail des points cités précédemment, on se reportera à [rev3, rev6, rev7, com16].

Cette structure a été mise à l'épreuve en y incorporant les données venant des formations Simuligne et Copéas, puis de celles concernant des formations à partenariat international

impliquant des interactions multilatérales : Ecofralin, Infral et Tridem06, particulièrement avec des enjeux interculturels [com19, com20, com22, ouv3, rev1, com1, com2, com3, com7, com8, com15].

Nous avons ensuite défini la notion de corpus distinguable.

Un corpus d'apprentissage correspondant à une expérience de formation est un objet de très grande taille. Mais les chercheurs s'intéressent souvent à des phénomènes plus limités. Évoluer au milieu d'un ensemble trop important de données peut alors devenir un obstacle.

Afin de travailler sur des unités intermédiaires, nous avons constitué des **corpus distinguables**. À partir d'un corpus d'apprentissage, que l'on qualifiera de **corpus global**, il est possible de produire des corpus distinguables, correspondant au grain habituellement retenu par un chercheur pour y accomplir une analyse sur un phénomène précis. Il peut se construire en segmentant le corpus global de deux façons très différentes : soit en ne retenant qu'une sous-partie du corpus global, soit en sélectionnant de façon longitudinale un ensemble de données (par exemple, tous les forums d'un groupe dans une formation, ou de tous les groupes).

Le corpus distinguable est tout à la fois un sous-corpus du corpus d'apprentissage et un corpus en soi. Son container est de même format que celui d'un corpus global. Mais à la différence de ce dernier, il est facilement téléchargeable sur un ordinateur personnel. Le chercheur dispose alors d'un ensemble comportant une description structurée du corpus contextualisé par rapport au corpus global (sous forme de commentaires libres et d'index précis renvoyant sur chacune des sous-parties du corpus global), et d'un ensemble de données soit prêtes à l'analyse, soit contenant déjà des résultats d'analyse. Enfin, le corpus distinguable s'insère dans un réseau inter-corpus. Il contient, par définition un ensemble de liens vers le corpus global (comme nous l'avons dit), mais aussi souvent des liens vers d'autres corpus distinguables issus du même corpus global. Ces liens sont des invitations à mener des analyses inter-corpus.

Toutes ces données ainsi rassemblées en corpus nous ont permis de développer de nouvelles analyses, en particulier sur les interactions multimodales et de poursuivre les publications sur ces sujets.

Notre recherche a donc été étroitement associée au développement des corpus. On le constatera en particulier en examinant les objectifs associés aux 3 types de corpus distinguables :

- associer publication scientifique et données (type 1) ;
- rassembler des données prêtes à l'analyse avec la mise en forme pour des outils/logiciels libres (type 2) ;
- partager des analyses avec des outils associés (type 3).

Pour le type 1, d'anciennes analyses et publications associées ont été reprises, réorganisées, renseignées et intégrées dans les corpus. L'idée étant de développer dans nos domaines de recherche la soumission d'articles de recherche avec le dépôt des données associées, cf. la notion d'ensemble de données pour la réplique référencée ci-dessus.

Les publications en rapport sont [rev5, rev6, com4, com6, com9, com10, com11].

Pour les corpus de type 2 et 3, nous avons travaillé avec les outils développés par les chercheurs des projets Calico (analyse de forums : <http://wims.crashdump.net/www/calico/>) et Tatiana

(analyse d'interactions multimodales : <http://lead.emse.fr/Download/tatiana.html>) et échangé avec ces chercheurs, qui ont systématiquement cités Mulce dans leurs communications.

Les publications en rapport sont [rev2, rev4, com5, com12, com13, com14, com17, com18, com21, com23, ouv1, ouv2].

4. Appréciation du degré de réalisation des objectifs initiaux, des obstacles rencontrés, des travaux qui restent à conduire

Tous les objectifs essentiels du projet Mulce ont été tenus et même au delà (voir section D). Au delà, car nous avons en plus :

- développé un site de documentation sur notre méthodologie permettant de communiquer nos livrables (Mulce.org, 2011) ;
- transféré la banque de corpus Mulce à Clermont pour assurer sa pérennisation (Mulce-databank, 2011) ;
- ajouté aux côtés des corpus d'apprentissage évoqués à l'origine (Simuligne, Copéas) les corpus d'apprentissage d'autres formations multinationales en ligne (Infral, Ecofralin, Tridem06) (Mulce-contenu, 2011).

Les tâches qui n'ont pas été réalisées concernent quelques analyses sur les données de nos corpus (abandon du partenaire LIP6) et le développement d'une procédure de dépôt en ligne de corpus (abandon d'un doctorat, maladie du premier ingénieur embauché). Les dépôts de nouveaux corpus doivent donc passer par l'intervention d'un administrateur de la base.

Les premiers travaux à conduire consisteraient à augmenter le nombre de corpus produits à partir de situations d'apprentissage, menées par des chercheurs hors Mulce. À ce jour, il n'existe dans notre banque qu'un seul corpus en ce sens : (Mulce-corpus-VMT, 2009).

5. Perspectives ouvertes par les résultats

Ces travaux permettent d'envisager de conduire des analyses impliquant plusieurs corpus, d'ouvrir la voie pour des analyses contradictoires ou complémentaires. La disponibilité de grands volumes d'interaction authentiques et caractérisés permettra de développer des analyses automatiques de type TAL (traitement automatique du langage), des tests de méthodes ou le calibrage d'outils (benchmarks), le dépassement du stade d'analyses n'exposant que des impressions à partir de données qualitatives réduites vers des analyses traitant des données contextualisées d'une large couverture à partir desquelles, des généralisations pourraient s'envisager.

Nous espérons voir se développer la soumission devant des comités scientifiques de relecture d'articles associant le dépôt de données qui feraient également l'objet d'une évaluation.

Enfin, dans l'esprit des recommandations de l'ANR et l'AERES, nous espérons contribuer à faire du développement de corpus (partageable), une activité de recherche reconnue au même niveau que les publications de grande qualité.

6. Appréciation de l'impact (scientifique, social, culturel, économique...)

Nous avons rassemblé des chercheurs de différentes disciplines travaillant sur les corpus, interactions, outils d'analyse d'interactions pour qu'ils constatent et/ou apprécient la complémentarité de leurs approches. Nos données ont commencé à être utilisées par des chercheurs ayant développé des outils d'analyse. À l'inverse, nous avons mis en forme des données d'autres chercheurs pour les intégrer à notre banque de corpus. Nos propos concernant association de publications et données ayant permis ces écrits sont écoutés, de même que la nécessité de prendre en compte le contexte des situations d'apprentissage et de structurer ces données pour mener des recherches plus systématiques. Mais les idées développées par Mulce étant nouvelles dans ces milieux, il est difficile d'en mesurer les impacts à court terme, l'évolution des méthodologies de recherche étant par nature très lente.

Pour donner un certain éclairage sur l'impact de nos travaux dans la communauté scientifique, voici la liste des *articles* (trouvés par Google Scholar) dont les auteurs ne sont pas dans Mulce et qui citent Mulce :

- GIGUET, E., LUCAS, N., BLONDEL, F.-M., & BRUILLARD, É. (2009). Share and explore discussion forum objects on the Calico website. In A. Dimitracopoulou, C. O'Malley, D. Suthers & P. Reimann (Eds.) Computer Supported Collaborative Learning Practices: CSCL 09 Community Events Proceedings (pp. 174-176). International Society of the Learning Sciences (ISLS).
- Caviale, O. « Faire parler » des échanges en ligne par une analyse tridimensionnelle », résumé à EPAL (Echanger pour Apprendre en Ligne) Grenoble, 2009.
- Teutsch, P., Bangou, F., Dejean-Thircuir, C. Faciliter l'accès aux échanges en ligne et leur analyse, le cas de ViCoDiLi. Numéro spécial : ÉPAL de la Revue Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation, Vol.15, 2008.
- Martínez-Monés, A., Marcos-García, J.A., Rodríguez-Triana, M. J. Fernández-San Miguel, D. Proposals for adaptive and interoperable IA systems. CSCL'2009 Workshop "Interaction Analysis and Visualization for Asynchronous Communication", Rhodes(Greece), 9-13 June 2009.
- Dyke, G., Lund, K., Girardot, J.-J. Tatiana, un environnement d'aide à l'analyse de traces d'interactions humaines. TSI 2010.
- Avouris, N., Kahrmanis, G., Fiotakis, G., Voyiatzaki, E. , Margaritis, M. Tools supporting Collaborative Learning and Interaction Analysis: Synergo and ActivityLens. EPAL Workshop, Grenoble, 2007.
- Boydens, I. « Entretien », Distances et savoirs 3/2009 (Vol. 7), p. 479-500. URL : www.cairn.info/revue-distances-et-savoirs-2009-3-page-479.htm.
- Martínez-Monés, A., Harrer, A., Dimitriadis, Y. An interaction-aware design process for the integration of interaction analysis into mainstream CSCL practices. Chapter in Sadhana Puntanbekar, Gijsbert Erkens, and Cindy Hmelo-Silver (Eds.). Analyzing interactions in CSCL: Methodologies, Approaches and Issues. Springer.

7. Pour les projets impliquant la constitution d'un corpus ou de base de données préciser les modalités de création, d'accès et les conditions de pérennisation

Modalités de création

Construction d'un corpus d'apprentissage par ordre chronologique (des procédures allégées pour des corpus moins structurés existent mais ne sont pas développées ici) :

- Collecte des données primaires pendant expérimentation de formation ;
- Constitution du fichier des acteurs avec codes, informations ethnographiques ; séparation licence privée (contrats de consentement éclairé signés par les participants) et publique ;
- Anonymisation des données primaires avec utilisation des codes d'acteurs ;
- Transformation des formats des données primaires en formats pérennes (élimination des formats propriétaires pour audio et vidéo, courriels, forum, etc.) ;
- Transcription (le cas échéant, facultatif) des données audio et vidéo correspondant à des interactions multimodales suivant code de transcription Mulce ;
- Description structurée (suivant schéma XML IMS-LD, incluant les métadonnées) du scénario pédagogique, du protocole de recherche avec liens vers les documents primaires : consignes pédagogiques, grilles d'entretiens, questionnaires, etc. ;
- Transformation des données d'instanciation, transcription comprises, dans le format XML Mulce-struct, incluant les métadonnées ;
- Intégration des différentes parties précédentes dans la structure globale du corpus d'apprentissage, le manifeste (schéma XML, IMS-CP) ;
- Élaboration des métadonnées générales du corpus (d'autres métadonnées existent dans chaque partie du manifeste) suivant le schéma OLAC ; mise à jour de la déclaration OLAC ;
- Dépôt du corpus dans la banque Mulce ;

Peuvent ensuite être élaborés les corpus distinguables.

Le dépôt d'un nouveau corpus (LETEC ou corpus distinguable : analyse) dans la banque Mulce doit passer par les administrateurs de la plateforme (cf. section 6).

Modalités d'accès et d'utilisation

L'accès à tous les corpus et objets de la banque Mulce est libre (*open access*).

L'accès aux métadonnées peut se faire à travers le serveur OLAC pour la base entière ou corpus par corpus : voir (OLAC-Mulce, 2011).

L'accès à la liste des corpus et à leur fiche descriptive est ouvert à tout internaute sans aucune inscription préalable : voir (Mulce-databank, 2011)

Seuls le parcours et le téléchargement des corpus nécessitent l'authentification de l'utilisateur et son acceptation de la licence d'utilisation (Mulce-licence, 2010) sur la plateforme Mulce. À ce jour, la création d'un compte sur la plateforme Mulce est ouverte à tous et ne nécessite aucune intervention des administrateurs.

Rappelons que dans la partie licence de nos corpus figurent les formulaires de contrats de consentement éclairé signés par les participants (Mulce-ethique, 2010). Les contrats eux-mêmes n'étant pas anonymes sont conservés par l'éditeur des corpus.

Condition de pérennisation

Le transfert de la banque Mulce, développé au départ chez le partenaire 2 (LIFC, Laboratoire d'Informatique), vers le partenaire 1 (Laboratoire de Recherche sur le Langage , LRL) vient d'avoir lieu début 2011 : voir (Mulce-databank, 2011). Ce transfert a fait l'objet d'une coopération de la MSH (Maison des Sciences de l'Homme) de Clermont-Ferrand qui fournit les machines serveurs, systèmes de sauvegarde et sécurité. Le LRL dispose d'un ingénieur informatique. Par ailleurs T. Chanier en est devenu le directeur.

Ce transfert garantit une certaine longévité à la banque Mulce car elle quitte une unité de recherche qui n'a plus d'axe de recherche en relation avec les sciences humaines pour intégrer une unité en sciences du langage, partie intégrante de la MSH (CNRS : Unité Mixte de Service, en phase de reconnaissance comme Unité de Recherche et Service).

Une pérennisation à plus long terme impliquerait une coopération avec le TGE-Adonis.

8. Exploitation et dissémination des résultats

Les sections C.1 et C.2 concernant les publications et les actions de diffusion et de valorisation de l'équipe Mulce indiquent les moyens utilisés pour toucher différentes communautés scientifiques (EIAH, sciences du langage, CSCL, TICE) par des écrits, communications, organisations de ou participations à des séminaires, symposiums, ateliers aux niveaux national et international.

L'équipe Mulce a en particulier organisé un symposium national dans le cadre de la conférence Épal en 2007, puis l'atelier "*Dissemination and comparison of research findings : developing Learning and Teaching Corpora (LETEC)*" dans le cadre de la conférence européenne Eurocall en 2010. Elle y a invité des chercheurs connus internationalement et travaillant sur des corpus dans les communautés apprentissage des langues et CSCL : voir le programme à (Eurocall-Mulce, 2010).

Citons encore ces éléments non mentionnés ailleurs :

- Le dépôt des forums de Simuligne dans Calico a permis une analyse (lexicale) complémentaire de ces forums sur la question de la cohésion des groupes dans Simuligne (papier court à CSCL 2011) ;
- le cluster ISLE (« Personnalisation des EIAH ») a lancé une thèse sur la question des corpus de traces et du benchmarking des outils dans le domaine des EIAH.

C. Publications, diffusion et valorisation

C.1 - Liste des publications et communications

Avertissement : toutes les publications des participants Mulce ont été déposées dans l'archive Edutice et y sont donc librement accessibles. : <http://edutice.archives-ouvertes.fr/>

Revue à comité de lecture

Monopartenaires

- [rev1] Audras, I., Chanier, T. (2008). "Observation de la construction d'une compétence interculturelle dans des groupes exolingues en ligne". *Apprentissage des Langues et Système d'Information et de Communication (ALSIC)*, vol. 11 (1). pp 175-204 <http://alsic.revues.org/index865.html>
- [rev2] Chanier, T. (2007) "Commentary: Open Access to Research and the Individual Responsibility of Researchers". *Language Learning & Technology*, vol. 11, 2 (2007). pp 142-148 . <http://llt.msu.edu/vol11num2/pdf/chanier.pdf>
- [rev3] Chanier, T., Ciekanski, M. (2010). "Utilité du partage des corpus pour l'analyse des interactions en ligne en situation d'apprentissage : un exemple d'approche méthodologique autour d'une base de corpus d'apprentissage". *Apprentissage des Langues et Systèmes d'Information et de Communication (ALSIC)*, 13 <http://alsic.revues.org/index1666.html>
- [rev4] Reffay, C. (2009). "L'accès libre aux revues, articles et données de recherche : retour sur la première conférence de l'OASPA". Rubrique dans la revue *Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation*, Vol.16, 2009. pp. 11-21.
- [rev5] Ciekanski, M., Chanier, T (2008). "Developing online multimodal verbal communication to enhance the writing process in an audio-graphic conferencing environment". *Recall*, vol. 20 (2), Cambridge University Press. 162-182. doi:10.1017/S0958344008000426

Multipartenaires

- [rev6] Betbeder, M.-L., Ciekanski, M., Greffier, F., Reffay, C., Chanier, T. (2008). "Interactions multimodales synchrones issues de formations en ligne : problématiques, méthodologie et analyses". *Revue Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation*, Vol.15, 2008. 29 p.
- [rev7] Reffay, C., Chanier, T., Noras, M., Betbeder, M.-L. (2008). "Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche". *Revue Sciences et Technologies de l'Information et de la Communication pour l'Education et la Formation*, Vol.15, 34 p. 2008.

Communications

Monopartenaires

- [com1] Audras, I., Chanier, T. (2007). "Intercultural experience and interactions in multilingual online groups". EUROCALL 2007, *Mastering Multimedia: Teaching Languages Through Technology*, University of Ulster, Coleraine Campus, les 5-8 septembre 2007.
- [com2] Audras, I. Chanier, T. (2007). "Observation de la construction d'une compétence interculturelle dans des groupes exolingues en ligne". In Lamy, Mangenot, Nissen (coord., 2007) *Actes du colloque Échanger pour apprendre en ligne (ÉPAL)*. Grenoble, 7-9 juin 2007. <http://w3.u-grenoble3.fr/epal/actes.html>

- [com3] Audras, I., Chanier, T. (2006). "Tridém et interaction à l'oral et à l'écrit dans une formation à distance en langue". *Actes de la conférence TICE 2006 (Technologies de l'Information et de la Communication pour l'Enseignement supérieur et l'Entreprise 2006)*, Toulouse, octobre. <http://edutice.archives-ouvertes.fr/edutice-00087737>
- [com4] Betbeder, M.-L., Tissot, R., Reffay, C. (2007). "Recherche de patterns dans un corpus d'actions multimodales". *Actes de la conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH 2007)*, Lausanne, Suisse, pages 533--544, Juin 2007.
- [com5] Chanier, T. (2007). "Comment réaliser l'accès libre aux résultats de la recherche en éducation et formation ?", *Congrès AREF (Actualités de la Recherche en Education et Formation)*, Strasbourg, 28-31 août 2007." (invité).
- [com6] Ciekanski, M., Chanier, T., Tajariol, F. (2009). "Du partage des corpus à l'analyse des interactions multimodales en ligne dans des situations d'apprentissage (de langues) : quelle méthodologie pour la recherche sur corpus d'apprentissage ?". *Colloque Échanger pour apprendre en ligne (ÉPAL)*. Grenoble, juin 2009.
- [com7] Ciekanski, M., Chanier, T. (2009). "Critical events and face-work in Multilateral Online Exchanges for Language and Culture Learning". *EUROCALL 2009, New Trends in CALL: working together*, Universidad Politécnica de Valencia, Gandia, septembre 2009.
- [com8] Ciekanski, M., Chanier, T. (2008). "Designing telecollaboration for fostering meetings in intercultural learning context: towards a general framework beyond cultural variability?". *Conférence CALL 2008 "Practice-Based & Practice-Oriented CALL Research"*, Antwerp, 30 août - 2 septembre 2008.
- [com9] Ciekanski, M., Chanier, T. (2008). "Soutenir la production écrite en L2 dans un environnement audio-graphique synchrone: stratégies communicatives et stratégies d'apprentissage en contexte multimodal", *Colloque ACEDLE (Association des Chercheurs et Enseignants Didacticiens des Langues Étrangères)*, Université de Strasbourg 2, les 17-19 janvier 2008.
- [com10] Ciekanski M., Chanier, T. (2007). "L2 proficiency in audio-synchronous online environments: rethinking communicative competence in the framework of multimodal interactions", *BAAL Cup-Seminar 2007, Spoken Online Learning Events: The need for a new paradigm in languages research and practice*, Milton Keynes, les 22 et 23 Juin 2007.
- [com11] Ciekanski M., Chanier, T. (2007). "Writing competence and oral participation in audio-graphic conferencing environment for ESL learners: towards an integrative approach of linguistic skills in multimodal communication context". *EUROCALL 2007, Mastering Multimedia: Teaching Languages Through Technology*, University of Ulster, Coleraine Campus, les 5-8 septembre 2007.
- [com12] Noras, M., Reffay, C., Betbeder, M.-L. (2007). "Structuration de corpus de formation en ligne en vue de leur échange". In *Actes de la conférence Environnements Informatiques pour l'Apprentissage Humain (EIAH 2007)*, Lausanne, Suisse, pages 59-64, Juin 2007.
- [com13] Reffay, C., Betbeder, M.-L. (2009). "Sharing corpora and tools to improve interaction analysis". *In proceedings of the EC-TEL 2009, 4th European Conference on Technology Enhanced Learning, Learning in the Synergy of Multiple Disciplines LNCS 5794, Springer, Nice, France, pages 196-210, October 2009.*
- [com14] Reffay, C., Betbeder, M.-L. (2009). "Extending validation of tools and analyses in CSCL situations: How to collaborate on interaction analysis?". *Pre-conference Workshop on Interaction analysis and visualization for asynchronous communication. Computer Supported Collaborative Learning (CSCL 2009)*, Rhodos, Greece, June 2009.

- [com15] Reffay, C., Betbeder, M.-L. (2009). "Construction of a Mulce corpus from the VMT dataset". Pre-conference Workshop on Common Objects for Productive Multivocality in Analysis. Computer Supported Collaborative Learning (CSCL 2009), Rhodos, Greece, June 2009.
- [com16] Teutsch, P, Piat, F, Reffay, C. (2009). "Anonymizing and sharing corpora of online training courses". CSCL'2009 Workshop "Interaction Analysis and Visualization for Asynchronous Communication", Rhodos, Greece, 9-13 June 2009.

Multipartenaires

- [com17] Betbeder, M.-L., Ciekanski, M., Greffier, F., Reffay, C., Chanier, T. (2007). "Comment spécifier, codifier et représenter les interactions multimodales synchrones issues de formations en ligne : problématiques, méthodologie et analyses". In Lamy, Mangenot, Nissen (coord., 2007) *Actes du colloque Échanger pour apprendre en ligne (ÉPAL)*. Grenoble, 7-9 juin 2007.
- [com18] Reffay, C., Chanier, T., Noras, M., Betbeder, M.-L. "Contribution à la structuration de corpus d'apprentissage pour un meilleur partage en recherche", In Lamy, Mangenot, Nissen (coord., 2007) *Actes du colloque Échanger pour apprendre en ligne (ÉPAL)*. Grenoble, 7-9 juin 2007.

Avec partenaires internationaux

- [com19] Abendroth-Timmer, D., Chanier, T., Ciekanski, M., Bechtel, M., Henning, E-V. (2010). "Du développement à l'investigation de la compétence interculturelle en pratique à partir des interactions à l'oral et à l'écrit dans des échanges en ligne à distance." *Colloque "Plurilingualism and Pluriculturalism in a Globalised World: which Pedagogy?" (PLIDAM)*, 17-19 Juin, Paris.
- [com20] Abendroth-Timmer, D., Bechtel, M., Chanier, T., Ciekanski, M. (2009). "From developing to investigating intercultural competence in practice through oral and written interactions in online exchanges", *Kongress für Fremdsprachendidaktik der Deutschen Gesellschaft für Fremdsprachenforschung (DGFF-Tagung)*, Universität Leipzig, octobre 2009.
- [com21] Lamy, M-N., Klarskov Mortensen, H. J. (2010). "Using concordance programs in the Modern Foreign Languages classroom". Module 2.4 in Davies G. (ed.) *Information and Communications Technology for Language Teachers (ICT4LT)*, Slough, Thames Valley University [Online]. Available from: http://www.ict4lt.org/en/en_mod2-4.htm
- [com22] López, M.T., Chanier, T. (2008). "Corpus de Aprendizaje: Conformación, Estructura y Construcción de un Corpus Virtual a partir de La Experiencia Ecofralin". *Comunicación en el encuentro internacional de investigación en educación. Institución Universitaria Iberoamericana*, 12, 13 y 14 de Noviembre 2008, Bogotá, Colombia.
- [com23] Reffay, C., Teplovs, C., Blondel, F.-M. (2011). "Productive re-use of CSCL data and analytic tools to provide a new perspective on group cohesion". Short paper accepted for the International conference on Computer Supported Collaborative Learning, Hong-Kong, July 2011.

Ouvrages ou chapitres d'ouvrages

Monopartenaires

- [ouv1] Reffay, C., Betbeder, M.-L., Dyke, G. "Data sharing in CSCR: towards in-depth long term collaboration". Chapter in book "Collaborative and Distributed E-Research: Innovations in Technologies, Strategies and Applications", IGI Publishing. (En préparation)

Multipartenaires

- [ouv2] Reffay, C., Chanier, T., Avouris, N., Romary, L. *Actes du Symposium "Corpus d'apprentissage en ligne : Conception, réutilisation, échange"*, intégré au colloque ÉPAL (Échanger Pour Apprendre

en ligne) <http://edutice.archives-ouvertes.fr/edutice-00161113/fr/>

Avec partenaires internationaux

[ouv3] Lewis, T. Chanier, T. & Youngs, B. (2011) (dir.) "*Multilateral Online Exchanges for Language and Culture Learning*". Numéro thématique de la revue *Language Learning & Technology (LLT)*, vol. 15, 1 <http://llt.msu.edu/issues/february2011/>

Tableau récapitulatif des publications

Publications				
	Articles acceptés dans des revues à comité de lecture	Ouvrages ou chapitres d'ouvrages	Communications	Publications soumises ou en préparation
monopartenaires	5	1	16	
multipartenaires	2	1	2	
Avec partenaires internationaux		1	5	1 chapitre

C.2 – Actions de diffusion et de valorisation

- **Développement d'un nouveau partenariat ou de nouvelles collaborations**
 - Gerry Stahl : VMT
 - Eric Bruillard : Calico
 - Elisabeth Delozanne : Corpus d'algèbre (avec Sésamath)
- **Participation à des activités d'expertise (collectives ou individuelles), de conseil, de formation**
 - Modification de la maquette du master Linguistique-Informatique de l'université Blaise Pascal (2012-2015) de façon à intégrer la notion de corpus d'apprentissage LETEC (aspects théoriques et méthodologiques)
- **Participation à des activités de vulgarisation : Séminaires, ateliers, posters.**

L'action principale qui a mobilisé toute l'équipe Mulce est l'organisation, sous la direction de Marie-Laure Betbeder et Maud Ciekanski, du Symposium "*Dissemination and comparison of research findings: Developing Contextualized Learning and Teaching Corpora (LETEC)*", en septembre 2010, lors de la conférence internationale EUROCALL à Bordeaux. A cette occasion, nous avons invité 5 chercheurs étrangers et autant de chercheurs français de différentes communautés (CSCL, apprentissage des langues) travaillant sur les corpus, les outils d'analyse des interactions. Voir Eurocal-Mulce (2010).

- Blondel, F.-M., Degache, C., Giguët, E., Reffay, C. (2011). Analyses de discussions en ligne avec Calico. *Organisation de l'atelier (ouvert) à EPAL (2011)*, Grenoble, 23 juin.
- Reffay, C. (2010). Partage d'outils et de données de recherche : des concepts à la mise en œuvre. *Séminaire AIDA*, UPMC, Paris-6, 5 nov. 2010 [<http://lutes.upmc.fr/aida/Reunions/05-11-2010.html>]
- Reffay, C. (2010). Partage de corpus et d'outils pour améliorer l'analyse des interactions : quelques résultats de Mulce. *Séminaire Cluster ISLE*, Chambéry, 21 sept. 2010.

- Reffay C. and Betbeder M.-L. (2009). Construction of a Mulce corpus from the VMT dataset Pre-conference *Workshop on Common Objects for Productive Multivocality in Analysis. Computer Supported Collaborative Learning (CSCL 2009)*, Rhodos, Greece, June 2009. [<http://edutice.archives-ouvertes.fr/edutice-00410225/fr/>]
- Reffay C. and Betbeder M.-L. (2009). Improving CSCL indicators by sharing multimodal teaching and learning Corpora. *Poster in CSCL 2009, Computer Supported Collaborative Learning conference*, Rhodos, Greece, June 2009. [<http://edutice.archives-ouvertes.fr/edutice-00410226/fr/>]
- Reffay, C. & Bruillard, E. (2008). Traces d'interaction en ligne : vers un partage des corpus, outils et analyses. *Journée annuelle du GDR I3*, organisé par P. Tchounikine, Lille, 14 mai 2008.
- Reffay, C. (2007). Corpus d'apprentissage : description, structuration, échange et capitalisation des données et des analyses. *Séminaire LIRIS*, PPF Apprentice, Lyon, juillet 2007.
- Reffay, C. & Mille, A. (2011). Partager des données d'observation pour la recherche en EIAH : traces d'activité d'apprentissage. *Organisation de l'atelier (ouvert) à EIAH2011*, Mons, Belgique. 24 mai. 6 soumissions, 5 acceptées (dont une de PSLC DataShop) [<http://www.stef.ens-cachan.fr/eiah2011/>]

Implications dans les formations

- Master M 2 mention recherche : didactique des sciences et techniques, ENS-Cachan. Depuis Septembre 2010, les étudiants (qui font traditionnellement une réplique d'analyses en partant d'un article extérieur au laboratoire STEF) doivent déposer les données qu'ils recueillent durant leur projet tuteuré de méthodologie en recherche (réplique). Voir la fiche de dépôt à remplir : Mulce-dépôt (2011).
- **Lancement de produit ou service,...**
 - plate-forme Mulce-databank (2011) : la banque de corpus Mulce ;
 - site de documentation Mulce.org (2011) : le site exposant la méthodologie Mulce et donnant accès aux livrables ;
 - Ouverture dans la MSH de Clermont-Ferrand de serveurs diffusant les résultats des travaux des unités de recherche associées avec en premier Mulce-databank (2011) : voir le projet de la MSH pour l'évaluation de l'AERES.
- **Actions de normalisation**
 - Déclaration à OLAC d'un nouveau type de corpus (Olac-Letec,2010) et de la banque de corpus Mulce (Olac-Mulce, 2011) ;
 - Protocoles concernant l'éthique et les droits : voir Mulce-ethique (2010).
- **Brevets nationaux et internationaux, licences et autres éléments de propriété intellectuelle consécutifs au projet**
 - N'ayant pu trouver en France d'endroit où déposer des résultats et données de recherche à la manière de l'archive *Dataverse (thedata.org/)*, et afin d'exprimer, avant l'ouverture de notre serveur, notre propriété intellectuelle sur la structure Mulce (Mulce-struct, 2010) et la notion de corpus d'apprentissage LETEC, nous avons décidé de faire le dépôt du premier corpus dans l'archive Cogprint (2009) ;
 - Licence d'utilisation des corpus Mulce, voir Mulce-licence (2010).

- **Autres**

- L'intégration (par détachement) de C. Reffay à l'unité de recherche STEF s'est faite sur la question du partage des données de recherche. Un projet de partage d'une collection de bulletins d'associations de spécialistes (enseignement de la technologie) est démarré au laboratoire STEF. Une réflexion est aussi en cours sur les modalités de partage de données de recherche en éducation en abordant la question par type de données recueillies (Questionnaires, entretiens, vidéos, Archives, etc.)
- T. Chanier, devenu directeur de l'UR LRL a pu impulser la mise au format corpus d'apprentissage des données recueillies dans les nouveaux projets d'apprentissage à distance et en ligne.

Tableau récapitulatif des actions de diffusion/valorisation

Diffusion/valorisation			
	Autres articles (dans revues sans comité de lecture,...)	Conférences	Autres (expositions, films, outils, plate-forme technique,...)
Monopartenaires		- Organisation d'ateliers / workshop : 2 - Séminaires / journées : 4 - Poster : 1	- Formations : 2
Multipartenaires		Organisation d'ateliers / workshop : 1	- Plates-formes : 2 - Normalisation : 2
Avec partenaires internationaux		Organisation d'ateliers / workshop : 1 –(idem multipartenaire)	

C.4 – Références Mulce

- Cogprint (2009). Chanier, Thierry and Lamy, Marie-Noelle and Reffay, Christophe and Betbeder, Marie-Laure and Ciekanski, Maud (2009) *LETEC (Learning and Teaching Corpus) Simuligne*. <http://cogprints.org/6431/>
- Eurocall-Mulce (2010). *Workshop "Dissemination and comparison of research findings: Developing Contextualized Learning and Teaching Corpora (LETEC)"*, September 8th 2010, EUROCALL Conference Bordeaux [<http://ubpweb.univ-bpclermont.fr/HEBERGES/mulce/spip.php?article9> ; <http://ubpweb.univ-bpclermont.fr/HEBERGES/mulce/spip.php?article29>]
- Olac-Mulce (2011) *Description des corpus Mulce au sein de OLAC* [<http://www.language-archives.org/archive/mulce.org> ; http://www.language-archives.org/archive_records/mulce.org]
- Olac-Letec (2010). *Définition d'un nouveau type de corpus (Letec) au sein de Olac* [http://lrl-diffusion.univ-bpclermont.fr/mulce/metadata/mce_LETECorpus-en.pdf ; http://lrl-diffusion.univ-bpclermont.fr/mulce/metadata/mce-schemas/mce_olac_letec.xsd]
- Mulce-contenu (2011). *Vue d'ensemble sur le contenu de la banque de corpus Mulce*. Mulce.org [<http://ubpweb.univ-bpclermont.fr/HEBERGES/mulce/spip.php?article30>]
- Mulce-corpus-VMT (2009). *LETEC based on an experience in mathematics, the Virtual Math Team (2006)* [http://mulce.univ-bpclermont.fr:8080/PlateFormeMulce/VIEW/PUBLIC/03/VMeta.do?adr=VMT/Corpus_objets/mce-vmt-letec-teamc]
- Mulce-databank (2011). Banque de corpus Mulce. Université Blaise pascal, MSH [<http://mulce.univ-bpclermont.fr:8080/PlateFormeMulce/>]
- Mulce-dépôt (2011). Dépôt de corpus (procédure simplifiée). Mulce.org [<http://ubpweb.univ-bpclermont.fr/HEBERGES/mulce/spip.php?article60>]

Mulce-ethique (2010). Protocoles éthique et droits avec documents associés. Mulce.org [<http://ubpweb.univ-bpclermont.fr/HEBERGES/mulce/spip.php?article38>]

Mulce-licence (2010). Licence d'utilisation des corpus Mulce. Mulce.org [http://lrl-diffusion.univ-bpclermont.fr/mulce/metadata/vdex/mce_licence.xml]

Mulce.org (2011). *Site de documentation sur la méthodologie Mulce* [<http://mulce.org>]

Mulce-struct (2010). *Schéma de la structure d'un corpus LETEC Mulce*. [<http://lrl-diffusion.univ-bpclermont.fr/mulce/metadata/mce-schemas/mce-letec-disting.xsd>]

Mulce-transcript (2009). *Conventions de transcriptions d'interactions multimodales* [<http://ubpweb.univ-bpclermont.fr/HEBERGES/mulce/spip.php?article45>]

D. Pour les projets partenariaux, liste des livrables et affectation éventuelle à chaque partenaire

Intitulé de recherche	Etapas	2008		2009		2010		Partenaires concernés
		S1	S2	S1	S2	S1	S2	
Tstruct1	Spécification Mulce-Struct corpus apprentissage	fait						Part 2
Tstruct2	modèle avec multimodalité			fait				Part 1
Tass1	assemblage d'un corpus en corp. apprentissage (Simuligne)			fait				Part 1 et 2
Tass2	assemblage d'un corpus multimodal (Copeas)				fait			Part1
Tmult1	Modèle transcriptions multimodales et transcriptions corpus		fait					Part1
Tmult2	Analyses multimodales et études outils d'aide à analyse				fait			Part1
Tpf1	Développement serveur Mulce : dépôt et consultation					fait		Part2
Tpf2 et Tana_v1	Intégration outils LIP6 et analyses avec outils LIP6	abandonné						LIP6
Tdoc	site explication methodo. (nouvelle tâche)						fait	Part1 et 2
Tana_v2	Analyses sur Tass1 et Tass2 à partir outils Part (idem Tmult2)				fait			Part1
Tdr1	contrats de cession des droits et éthique	fait						Part1
Tdr2	Protocole dépôt et consultation corpus (licences)				fait			Part2
Tcom1	Organisation et communication dans événements scientifiques				fait		fait	Part1 et 2
Tcom2	Symposium sur corpus d'apprentissage						fait	Part1 et 2
Tserv1 et 2	Achat, configuration, mise en place serveur		fait					Part2
Transfert	Transfert du serveur de Besançon à Clermont (non prévu initialement)						fait	Part1 et 2
Dépôts	Dépôts supplémentaires de corpus d'apprentissage (non prévu initialement)						fait	Part1 et 2

Documents : on retrouvera sur le site Mulce.org (<http://mulce.org>), les nombreux documents produits par les partenaires du projet Mulce, soit sous forme d'articles du site, soit en document attachés.

Réunions en présentiel du projet : mars 2007, juin 2009 (colloque Epal), mars 2010, septembre 2010 (symposium Eurocall). La très grande partie du travail se faisait à distance avec des listes de discussions et des plates-formes de mises en commun des fichiers et travaux.

E. Bilan et suivi des personnels recrutés en CDD (hors stagiaires)

Identification				Avant le recrutement sur le projet			Recrutement sur le projet				Après le projet				
Nom et prénom	Sexe H/F	Adresse email (1)	Date des dernières nouvelles	Dernier diplôme obtenu au moment du recrutement	Lieu d'études (France, UE, hors UE)	Expérience prof. Antérieure, y compris post-docs (ans)	Partenaire ayant embauché la personne	Poste dans le projet (2)	Durée missions (mois) (3)	Date de fin de mission sur le projet	Devenir professionnel (4)	Type d'employeur (5)	Type d'emploi (6)	Lien au projet ANR (7)	Valorisation expérience (8)
Chipeaux Sébastien	H	sebastien.chipeaux@lifc.univ-fcomte.fr , sebastien.chipeaux@laposte.net	1 ^{er} Février 2011	Master2 Pro Informatique	Besançon, France	1,5 an	LIFC	Ingénieur	9 mois	Nov. 2008	CDD	ERP	Ingénieur doctorant	Partenaire LIFC	expérience de la recherche
Qingrui Zeng	H	qingrui.zeng@hotmail.com		master	France		LIFC, mais travail au LRL	Ingénieur	3 mois	décembre 2010					
TAJARIOL FEDERICO	H			Doctorat sciences cognitives	France		LRL-Laseldi	Ingénieur	4 mois	février 2009	MCF	enseignement et recherche publique	enseignant-chercheur	oui	oui
TOMC Sandra	F			Doctorat sciences du langage	France	1	LRL-Laseldi	Ingénieur	3 mois	juillet 2008	MCF	enseignement et recherche publique	enseignant-chercheur	non	oui