



Thèse présentée
par

Loïc LIÉGEOIS

en vue de l'obtention du grade de Docteur de l'Université Blaise Pascal – Clermont
Université

Spécialité : Sciences du langage

**Usage des variables phonologiques dans un
corpus d'interactions naturelles parents-enfant :
impact du bain linguistique et dispositifs cognitifs
d'apprentissage.**

Volume 2 : Annexes

Membres du jury

Damien CHABANAL, Université Blaise Pascal, Clermont Université, Directeur

Thierry CHANIER, Université Blaise Pascal, Clermont Université, Directeur

Jean-Pierre CHEVROT, Université Stendhal Grenoble 3, Rapporteur

Harriet JISA, Université Lumière Lyon 2, Examinatrice

Christophe PARISSE, INSERM, Université Paris Ouest Nanterre La Défense, Rapporteur

Préparée au sein du Laboratoire de Recherche sur le Langage, EA999

sous la direction de Thierry CHANIER et Damien CHABANAL

Ecole doctorale Lettres, Sciences Humaines et Sociales, ED370

Soutenance le vendredi 7 novembre 2014

Sommaire

Sommaire	I
Annexe 1 : Appel à participation au projet ALIPE.....	1
Annexe 2 : Accord de consentement éclairé.....	2
Annexe 3 : Contextes d'interaction de chaque corpus distinguable.....	4
Annexe 4 : Programmes informatiques	12

Annexe 1 : Appel à participation au projet ALIPE

□

Loïc Liégeois
Docteur allocataire
06-XX-XX-XX-XX
XXX@hotmail.fr

Laboratoire de Recherche sur le Langage
Université Blaise Pascal, Clermont-
Ferrand

Appel à participation à un projet de recherche.

Madame, Monsieur,

Je débute cette année un doctorat en sciences du langage dans le but d'étudier l'acquisition du langage chez les enfants. Cette recherche sera financée par le ministère de la recherche et de l'enseignement supérieur.

Afin d'observer le langage des enfants, j'ai choisi de récolter des données en situation naturelle. J'ai donc pour objectif d'enregistrer (pendant une semaine, à raison d'une heure d'enregistrement par jour) les enfants à leur domicile à des moments où ils sont souvent en discussion avec leurs parents (repas, bain, jeux, lecture...). Les deux familles qui participeront au projet auront à disposition un enregistreur numérique que les parents déclencheront quand ils le souhaiteront. Je n'interviens donc en aucun cas au domicile des familles.

Afin de sélectionner les deux enfants qui participeront au projet, j'ai retenu plusieurs critères :

- L'âge : les enfants doivent être nés entre octobre 2007 et février 2008. De plus, les deux enfants sélectionnés ne doivent pas avoir plus de 20 jours d'écart.
- Le sexe : les enfants qui participeront au projet seront soit deux garçons, soit deux filles.
- L'école : les deux enfants qui seront enregistrés doivent être dans la même classe.

Je me permets de vous contacter via ce courrier car votre enfant correspond à l'ensemble de ces critères. Vous comprendrez que la nature des critères retenus fait que très peu d'enfants ont été présélectionnés. Si vous le souhaitez, je vous propose de vous rencontrer à l'école, à votre domicile ou bien par téléphone afin de vous expliquer le projet plus en détail.

Je vous remercie d'avance et vous prie d'agréer, Madame, Monsieur, mes salutations les meilleures.

Cordialement,

Loïc Liégeois

Annexe 2 : Accord de consentement éclairé

□



Projet ALIPE Accord de consentement éclairé

Nous vous invitons à participer, dans le cadre du projet ALIPE que nous conduisons, à une étude sur les modalités d'apprentissage de l'oral chez l'enfant pré-lecteur, l'impact du bain linguistique et les dispositifs cognitifs d'apprentissage. Celle-ci a pour objectif d'étudier l'acquisition du langage par le jeune enfant.

Dans le cadre de ce projet, nous recherchons des familles acceptant d'être enregistrées au cours de scènes de vie quotidiennes où les interactions entre parents et enfant sont favorisées (repas, bain, jeux, lecture...).

Les informations personnelles recueillies au cours des enregistrements seront gardées confidentielles et ne seront pas diffusées. Ces informations seront brouillées sur le signal audio de l'enregistrement et les transcriptions qui en résultent seront anonymisées. Seuls les prénoms seront conservés sur ces transcriptions. Ces données transformées (transcription, études) et rendues anonymes seront, à l'avenir, à disposition d'autres chercheurs sur internet par le biais d'une plateforme dédiée.

Si, malgré les soins apportés à l'anonymisation, subsistent dans ces données des informations susceptibles de vous identifier et que vous estimez cette situation indésirable, il vous suffira de nous le signaler pour que l'anonymisation soit améliorée ou les données retirées. De plus, vous gardez le droit de nous demander de retirer tout passage que vous jugez indésirable.

Si vous nous donnez aujourd'hui votre accord pour participer à ce projet, sachez que vous gardez le droit de vous rétracter à tout moment de cette étude, sans avoir à vous justifier.

L'équipe de recherche vous tiendra au courant de l'avancée de ses travaux et vous fera part des principaux résultats auxquels vous aurez accès. Nous vous remercions par avance pour votre participation à cette collecte de données dans le cadre de notre projet de recherche. Nous nous tenons à votre disposition pour toute requête ou question :

□

projet de recherche. Nous nous tenons à votre disposition pour toute requête ou question :

Damien Chabanal (Maître de conférence, UFR LLSH Blaise Pascal, LRL, axe DEOL) :
damien.chabanal@univ-bpclermont.fr / 06-XX-XX-XX-XX

J'ai lu et compris les informations contenues dans ce document. J'accepte de participer à ce projet de recherche en sachant que je peux me retirer à tout moment. J'ai reçu une copie de cet accord.

Nom et Prénom des participants :

Signature : _____ Date :

Signature : _____ Date :

Nom et prénom de chercheur :

Signature : _____ Date :

Nom et prénom du directeur de recherche :

Signature : _____ Date :

Annexe 3 : Contextes d'interaction de chaque corpus distinguable

Identifiant du corpus	Description des contextes d'interaction
ali-baptiste-101227-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France. Le garçon joue avec son père avec des outils pour enfant (vis, marteau...).
ali-baptiste-101227-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon prend son goûter en compagnie de ses parents.
ali-baptiste-101227-3	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon prend son goûter en compagnie de ses parents.
ali-baptiste-101228-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans le salon. Le garçon, sa mère et son père parlent de sujets divers et écoutent de la musique.
ali-baptiste-101228-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la salle de bain. Le garçon prend son bain en compagnie de son père.
ali-baptiste-101228-3	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France. Le garçon, sa mère et son père discutent de sujets variés. Le père essaie de réparer un jouet de son fils.
ali-baptiste-101228-4	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-101228-5	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la chambre du garçon. Le père lit des livres à son fils, et tous les deux discutent des images. Le père pose beaucoup de questions dans cette session d'enregistrement. Le garçon décrit les images qu'il voit.
ali-baptiste-101229-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France. Le garçon est en train de dessiner. Il y a un assez long monologue du garçon dans cette session d'enregistrement.
ali-baptiste-101229-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon prend son goûter en compagnie de son père.

Identifiant du corpus	Description des contextes d'interaction
ali-baptiste-101230-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France. Le garçon, sa mère et son père parlent de sujets divers.
ali-baptiste-101230-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la salle de bain. Le garçon prend son bain en compagnie de son père.
ali-baptiste-101230-3	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la chambre du garçon et dans la cuisine. Le père aide son fils à s'habiller et les parents discutent du menu du réveillon.
ali-baptiste-101230-4	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France. La famille se prépare à passer à table.
ali-baptiste-101230-5	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-101230-6	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-101231-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le père cuisine et discute avec son fils.
ali-baptiste-101231-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Les parents cuisinent et discutent avec leur fils. La mère a une conversation téléphonique.
ali-baptiste-101231-3	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-110101-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la salle de bain. Le garçon prend son bain en compagnie de son père.
ali-baptiste-110101-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France. Le père aide son fils à s'habiller.
ali-baptiste-110101-3	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France. Le père aide son fils à s'habiller, puis ils vont jouer avec des voitures miniatures.
ali-baptiste-110101-4	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France. Le père et la mère discutent avec leur fils avant d'aller le coucher.

Identifiant du corpus	Description des contextes d'interaction
ali-baptiste-110101-5	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Les parents discutent avec leur fils pendant qu'il prend son goûter.
ali-baptiste-110103	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-1100706	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés. Les parents ont une longue conversation à laquelle le garçon ne participe pas. Le garçon va ensuite jouer à un jeu de rôle : il joue le rôle d'un cuisinier.
ali-baptiste-1100707	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-1100709	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-1100710-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-1100710-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-1100712	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-1100713	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-1100717	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-baptiste-1100720	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.

Identifiant du corpus	Description des contextes d'interaction
ali-baptiste-1100801	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Le garçon, sa mère et son père sont en train de manger et discutent de sujets variés.
ali-salome-081117-1	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France et plus précisément dans le salon de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette font des activités variées : ils discutent, regardent des catalogues de jouets ensemble, regardent des photos...
ali-salome-081117-2	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France et plus précisément dans la salle de bain de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père fait prendre le bain à sa fillette. La mère n'intervient pas dans cette séance.
ali-salome-081118	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père et la fillette font des activités variées : ils discutent, font des dessins, jouent avec des figurines... Ensuite, la mère arrive et tous les trois vont jouer à un jeu d'adresse. Puis tous les trois vont se mettre à table pour prendre le dîner.
ali-salome-081119	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France et plus précisément dans la cuisine de la famille, puis dans la salle de bain. Il a été recueilli à l'aide d'un enregistreur numérique. Alors que le père termine son dîner, la mère et la fillette jouent et discutent ensemble. Ensuite, le père va emmener sa fille prendre son bain.
ali-salome-081120	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France. Il a été recueilli à l'aide d'un enregistreur numérique. Au départ, le père et la fillette jouent à un jeu d'adresse. Puis, le père et la mère vont se relayer à tour de rôle pour jouer à divers jeux avec leur fillette pendant que l'autre prépare l'arrivée d'invités attendus pour dîner.
ali-salome-081121	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France. Il a été recueilli à l'aide d'un enregistreur numérique. Après que la mère ait habillé la fillette et fait sa toilette, toutes les deux vont jouer à des jeux variés : elles discutent, jouent à un jeu d'adresse, jouent avec des barettes... Le père n'intervient que rarement dans cette séance et ne prend pas part aux activités.

Identifiant du corpus	Description des contextes d'interaction
ali-salome-090710	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France dans la salle de bain puis dans la chambre de la fillette. Il a été recueilli à l'aide d'un enregistreur numérique. La fillette prend son bain en compagnie de son père. Ensuite, le père et la fillette vont jouer ensemble à divers jeux, notamment un jeu d'adresse.
ali-salome-090712	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France et plus précisément dans la salle de bain puis dans la chambre de la fillette. La fillette prend son bain avec son père, puis le père va lire des histoires à sa fillette dans sa chambre.
ali-salome-090713	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France, majoritairement dans la salle de bain. La fillette prend son bain avec son père, puis le père, la mère et la fillette vont préparer leurs affaires pour leur départ en vacances.
ali-salome-090714	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France et plus précisément dans le salon de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette font des activités variées : ils discutent et jouent ensemble, notamment à des jeux de rôle.
ali-salome-090716	Cet enregistrement se déroule au domicile des parents à Montigny-aux-Amognes, France et plus précisément dans la salle de bain et la chambre de la fillette. La fillette prend son bain en compagnie de son père. Puis, elle va préparer les jouets qu'elle désire emmener en vacances avec sa mère. Ensuite, toutes les deux vont jouer à un jeu de rôle, durant lequel la fillette va jouer le rôle d'une cuisinière et d'une serveuse.
ali-prune-071120-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la chambre de l'enfant. Il a été recueilli à l'aide d'un enregistreur numérique. Le père et la fillette font des activités variées : ils discutent, regardent un livre ensemble, jouent avec des figurines... La fillette va également prendre part à différents jeux de rôle comme le médecin, la cuisinière et la serveuse d'un restaurant. La mère n'intervient que rarement dans cette séance et ne prend pas part aux activités.
ali-prune-071120-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés. Après le repas, le père et la fille jouent à divers jeux de rôles (papa et maman, serveuse d'un restaurant).

Identifiant du corpus	Description des contextes d'interaction
ali-prune-071121-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la chambre de la fillette puis dans la cuisine. Il a été recueilli à l'aide d'un enregistreur numérique. Au départ, le père et la fillette jouent ensemble et discutent de sujets variés, notamment de ce que la fillette a fait à l'école. Puis la mère arrive, et toute la famille passe à table. Les discussions vont porter sur des sujets variés. On note dans ce passage un nombre important de discussions entre le père et la mère dans lesquelles la fillette essaient de s'insérer.
ali-prune-071121-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la chambre de la fillette. Il a été recueilli à l'aide d'un enregistreur numérique. Le père et la fillette discutent et jouent à divers jeux de rôles (serveuse et cuisinier d'un restaurant, papa et maman).
ali-prune-071121-3	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine. Il a été recueilli à l'aide d'un enregistreur numérique. La mère et la fillette cuisinent ensemble.
ali-prune-071122	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la chambre de la fillette. Il a été recueilli à l'aide d'un enregistreur numérique. Le père et la fillette jouent ensemble à des jeux de rôle au cours desquels la fillette joue le rôle d'un cuisinier et d'un serveur dans un restaurant. Le père et sa fille vont également regarder des livres d'image ensemble, et le père va lire quelques histoires à sa fille.
ali-prune-071124	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, ma mère et la fillette prennent leur petit déjeuner ensemble tout en discutant. Ensuite, le père et la fillette vont regarder des livres d'images ensemble, et le père va lire quelques livres à sa fille.
ali-prune-071125	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père et la fillette prennent leur petit déjeuner ensemble tout en discutant.
ali-prune-080714-1	Cet enregistrement se déroule sur le lieu de vacances de la famille, à Montpellier, France. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette viennent de finir leur petit déjeuner et planifient leur journée.

Identifiant du corpus	Description des contextes d'interaction
ali-prune-080714-2	Cet enregistrement se déroule sur le lieu de vacances de la famille, à Montpellier, France. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, et la fillette terminent leur déjeuner. Ils jouent ensemble à se poser des devinettes et discutent de sujets variés. La mère n'intervient pas au cours de cette session d'enregistrement.
ali-prune-080715-1	Cet enregistrement se déroule au domicile des parents, à Clermont-Ferrand, France et plus précisément dans la chambre de la fillette. Il a été recueilli à l'aide d'un enregistreur numérique. La mère et la fillette discutent et jouent ensemble avec divers jouets. Elles vont également jouer à des jeux de rôle : la fillette joue le rôle d'une cuisinière puis d'un vétérinaire. Le père n'intervient pas au cours de cette session d'enregistrement.
ali-prune-080715-2	Cet enregistrement se déroule au domicile des parents, à Clermont-Ferrand, France et plus précisément dans la cuisine. Il a été recueilli à l'aide d'un enregistreur numérique. Le père et la fillette déjeunent ensemble et discutent de sujets variés. La mère n'intervient pas au cours de cette session d'enregistrement.
ali-prune-080716	Cet enregistrement se déroule au domicile des parents, à Clermont-Ferrand, France et plus précisément dans la cuisine. Il a été recueilli à l'aide d'un enregistreur numérique. La mère et la fillette discutent ensemble de sujets variés pendant que la fillette prend son goûter. La mère et la fillette chantent quelques comptines. Le père n'intervient pas au cours de cette session d'enregistrement.
ali-prune-091123-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père et la fillette sont à table et discutent de sujets variés. La mère n'intervient pas dans cette session d'enregistrement
ali-prune-091123-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés.
ali-prune-091123-3	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés. Dans cette session, le père et la mère ont une longue discussion au cours de laquelle leur fillette intervient rarement.

Identifiant du corpus	Description des contextes d'interaction
ali-prune-091124-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés.
ali-prune-091124-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés.
ali-prune-091125-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés. Dans cette session, le père et la mère ont une longue discussion au cours de laquelle leur fillette intervient rarement.
ali-prune-091125-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés.
ali-prune-091126	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés.
ali-prune-091127-1	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés.
ali-prune-091127-2	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés.
ali-prune-091128	Cet enregistrement se déroule au domicile des parents à Clermont-Ferrand, France et plus précisément dans la cuisine de la famille. Il a été recueilli à l'aide d'un enregistreur numérique. Le père, la mère et la fillette sont à table et discutent de sujets variés.

Annexe 4 : Programmes informatiques

Programme LiveCode de l'application de conversion du corpus structuré au format XML-ALIFE vers le format CHAT

-- Début du programme

-- Remise sur fond blanc de toutes les lignes du texte en phase de traitement. Partie du script appelée entre chaque étape de traitement dans le but de repérer d'éventuelles erreurs d'annotation.

```

on metBlanc
  put "Texte" into fieldSource
  put "Resultat" into fieldResultat

  put the number of lines of field fieldResultat into nbl
  repeat with x = 1 to nbl

    set the backgroundColor of line x of field fieldResultat to white

  end repeat
end metBlanc

```

-- Conversion des balises contenant les informations sur la liaison

```

on traiterCasLiaison

  put "Texte" into fieldSource
  put "Resultat" into fieldResultat

  put the number of lines of field fieldSource into nbl
  repeat with x = 1 to nbl
    put line x of field fieldSource into laligne
    repeat while matchText (laligne,"<L.>.*(<.)(<.)(<./>).*</L.>",cs,ca,cp,oo) = true
      if matchText (laligne,"(<L.>.*(<.)(<.)(<./>).*</L.>)",debut,cs,ca,cp,oo,fin) = true then
        put debut & \
          "[^ Syntctx=" & cs & \
            " expecCons=" & ca & \
            " realCons=" & cp & \
            " obliOpt=" & oo & "]" & \
          fin into laligne
        put laligne into line x of field fieldResultat
        set the backgroundColor of line x of field fieldResultat to yellow
      end if
    end repeat
  end repeat
end traiterCasLiaison

```

-- Conversion des balises contenant les informations sur la langue des énoncés produits

```
on traiterCasLangue
  put the number of lines of field "Resultat" into nbl
  repeat with x = 1 to nbl
    put line x of field "Resultat" into laligne
    repeat while matchText
      (laligne,"(<L.>.*<(eng|ita|spa)>\s(\S+)\s<\/(eng|ita|spa)>(.*<\/L.>)",debut,langue,texte,langue2,fin) = true
      if matchText
        (laligne,"(<L.>.*<(eng|ita|spa)>\s(\S+)\s<\/(eng|ita|spa)>(.*)",debut,langue,texte,langue2,fin) = true then
          put debut& \
            texte@"s:"&langue& \
            fin into laligne
          put laligne into line x of field "Resultat"
          set the backgroundcolor of line x of field "Resultat" to yellow
        end if
      end repeat
    end repeat
  end repeat
end traiterCasLangue
```

-- Conversion des balises contenant les informations sur les répétitions

```
on traiterCasRepetition
  put the number of lines of field "Resultat" into nbl
  repeat with x = 1 to nbl
    put line x of field "Resultat" into laligne
    repeat while matchText
      (laligne,"(<L.>.*<g(.+)>\s(.*)\s<\/g.+>(.*<\/L.>)",debut,repétition,texte,fin) = true
      if matchText (laligne,"(<L.>.*<g(.+)>\s(.*)\s<\/g.+>(.)",debut,repétition,texte,fin) = true then
        put debut& \
          "<"&texte&"> "& \
          "[x "&repétition& \
          "]"& \
          fin into laligne
        put laligne into line x of field "Resultat"
        set the backgroundcolor of line x of field "Resultat" to yellow
      end if
    end repeat
  end repeat
end traiterCasRepetition
```

-- Conversion des balises contenant les informations sur les transcriptions alternatives

```
on traiterCasAlternative
```

```

put the number of lines of field "Resultat" into nbl
repeat with x = 1 to nbl
  put line x of field "Resultat" into laligne
  repeat while matchText (laligne,"(<L.>.*<alt>\s(.*)\s<a>\s(.*)\s</a>
</alt>.*</L.>)",debut,texte1,texte2,fin) = true
    if matchText (laligne,"(<L.>.*<alt>\s(.*)\s<a>\s(.*)\s</a>
</alt>.*</L.>)",debut,texte1,texte2,fin) = true then
      put debut& \
        "<"& \
        texte1& \
        ">"& \
        "[=? "& \
        texte2&"]" & \
        fin into laligne
      put laligne into line x of field "Resultat"
      set the backgroundcolor of line x of field "Resultat" to yellow
    end if
  end repeat
end repeat
end traiterCasAlternative

```

<p>-- Conversion des balises contenant les informations sur les cas de chevauchement d'énoncés</p>

on traiterCasChevauchement

```

put the number of lines of field "Resultat" into nbl
repeat with x = 1 to nbl
  put line x of field "Resultat" into laligne
  repeat while matchText
(laligne,"(<L.>.*<ens(1|2)>(\s<.*>\s\[.+\])\s</ens(1|2)>(.*)",debut,ensdebut,texte,ensfin
,fin) = true
    if matchText
(laligne,"(<L.>.*<ens(1|2)>(\s<.*>\s\[.+\])\s</ens(1|2)>(.*)",debut,ensdebut,texte,ensfin
,fin) = true then
      if ensdebut = "1" then
        put debut& \
          texte& \
          "[>]"& \
          fin into laligne
        put laligne into line x of field "Resultat"
        set the backgroundcolor of line x of field "Resultat" to yellow
      end if
      if ensdebut = "2" then
        put debut& \
          texte& \
          "[<]"& \
          fin into laligne
        put laligne into line x of field "Resultat"
        set the backgroundcolor of line x of field "Resultat" to yellow
      end if
    end if
  end repeat
end repeat

put the number of lines of field "Resultat" into nbl

```

```

repeat with x = 1 to nbl
  put line x of field "Resultat" into laligne
  repeat while matchText
(laligne,"(<L.>.*<ens(1|2)>(.*</ens(1|2)>(.*)",debut,ensdebut,texte,ensfin,fin) = true
  if matchText
(laligne,"(<L.>.*<ens(1|2)>\s(.*)\s</ens.>(.*)",debut,ensdebut,texte,fin) = true then
    if ensdebut = "1" then
      put debut& \
        "<"& texte&\
        "> [>]"& \
        fin into laligne
      put laligne into line x of field "Resultat"
      set the backgroundcolor of line x of field "Resultat" to yellow
    end if
    if ensdebut = "2" then
      put debut& \
        "<"& texte&\
        "> [<]"& \
        fin into laligne
      put laligne into line x of field "Resultat"
      set the backgroundcolor of line x of field "Resultat" to yellow
    end if
  end if
end repeat
end repeat
end traiterCasChevauchement

```

-- Conversion des balises contenant les informations sur les modes de production des énoncés

```

on traitementCasVoix
  put the number of lines of field "Resultat" into nbl
  repeat with x = 1 to nbl
    put line x of field "Resultat" into laligne
    repeat while matchText
(laligne,"(.*)<(crie|pleure|chante|pleure|rit|chuchote|lit|gémît|baille|tousse|soupon)>\s*(0)\s
* </(crie|pleure|chante|pleure|rit|chuchote|lit|gémît|baille|tousse|soupon)>(.*",debut,voix,d
edans,voix2,fin) = true
    if matchText
(laligne,"(.*)<(crie|pleure|chante|pleure|rit|chuchote|lit|gémît|baille|tousse|soupon)>\s*(0)\s
</(crie|pleure|chante|pleure|rit|chuchote|lit|gémît|baille|tousse|soupon)>(.*",debut,voix,de
dans,voix2,fin) = true then
      put debut& \
        dedans& \
        " [=! "&voix&"]"& \
        fin \
        into laligne
      put laligne into line x of field "Resultat"
      set the backgroundcolor of line x of field "Resultat" to yellow
    end if
  end repeat
end repeat

  put the number of lines of field "Resultat" into nbl
  repeat with x = 1 to nbl
    put line x of field "Resultat" into laligne

```

```

    repeat while matchText
(laligne,"(*)<(crie|pleure|chante|pleure|rit|chuchote|lit|gémît|baille|tousse|soupire)>\s*(.*)\
s*</(crie|pleure|chante|pleure|rit|chuchote|lit|gémît|baille|tousse|soupire)>(.*)",debut,voix,
dedans,voix2,fin) = true
    if matchText
(laligne,"(*)<(crie|pleure|chante|pleure|rit|chuchote|lit|gémît|baille|tousse|soupire)>\s*(.*)\
s</(crie|pleure|chante|pleure|rit|chuchote|lit|gémît|baille|tousse|soupire)>(.*)",debut,voix,d
edans,voix2,fin) = true then
        put debut& \
            "<&dedans&">"& \
            "[=! "&voix&"]"& \
            fin \
            into laligne
        put laligne into line x of field "Resultat"
        set the backgroundcolor of line x of field "Resultat" to yellow
    end if
end repeat
end repeat

end traitementCasVoix

```

-- Conversion des balises contenant les informations sur les locuteurs

```

on remplacements1
    put field "Resultat" into txt
    put replaceText (txt,"<L1>","-MOT:") into txt
    put replaceText (txt,"<L2>","-FAT:") into txt
    put replaceText (txt,"<L3>","-CHI:") into txt
    put replaceText (txt,"\s+</L.>", "") into txt
    put replacertext (txt,"<AT\|>\s", "") into txt
    put txt into field "Resultat"
end remplacements1

```

-- Conversion des balises contenant les informations sur l'adresse du discours

```

on traitementAdresse
    put the number of lines of field "Resultat" into nbl
    repeat with x = 1 to nbl
        put line x of field "Resultat" into laligne
        repeat while matchText (laligne,".*(<A./>).*",ad) = true
            if matchText (laligne,"((-MOT|-FAT).*)(<AA/>)(.*)",debut,locuteur,ad,fin) = true then
                if locuteur = "-MOT" then
                    put debut& \
                        fin& \
                        "-cr-"& \
                        "%add: à FAT" \
                        into laligne
                    put laligne into line x of field "Resultat"
                    set the backgroundcolor of line x of field "Resultat" to yellow
                end if
                if locuteur = "-FAT" then
                    put debut& \

```

```

        fin& \
        "-cr-"& \
        "%add: à MOT" \
        into laligne
    put laligne into line x of field "Resultat"
    set the backgroundcolor of line x of field "Resultat" to yellow
end if
end if
if matchText (laligne,"((-MOT|-FAT).*)(<AE/>).(*)",debut,locuteur,ad,fin) = true then
    put debut& \
        fin& \
        "-cr-"& \
        "%add: à CHI" \
        into laligne
    put laligne into line x of field "Resultat"
    set the backgroundcolor of line x of field "Resultat" to yellow
end if
end repeat
end repeat
end traitementAdresse

```

-- Conversion des balises contenant les informations sur les formes spéciales

```

on traitementMotsSpeciaux
    put the number of lines of field "Resultat" into nbl
    repeat with x = 1 to nbl
        put line x of field "Resultat" into laligne
        repeat while matchText
            (laligne,"(*)<(c|f|i|o|l|>\s*(\S+)\s*</(c|f|i|o|l|>(.*)",debut,marque,dedans,marque2,fin) =
            true
            if matchText
                (laligne,"(*)<(c|f|i|o|l|>\s*(.*)\s*</(c|f|i|o|l|>(.*)",debut,marque,dedans,marque2,fin) = true
                then
                    put debut& \
                        dedans&"@"& \
                        marque& \
                        fin \
                        into laligne
                    put laligne into line x of field "Resultat"
                    set the backgroundcolor of line x of field "Resultat" to yellow
                end if
            end repeat
        end repeat
    end traitementMotsSpeciaux

```

-- Conversion des informations sur les marqueurs de début et de fin d'énoncé

```

on remplacement2
    put field "Resultat" into txt
    put replaceText (txt,"-cr-",cr) into txt

```



```

put replaceText (txt,"-MOT:\s+","*MOT:-tab-") into txt
put replaceText (txt,"-FAT:\s+","*FAT:-tab-") into txt
put replaceText (txt,"-CHI:\s+","*CHI:-tab-") into txt
put replaceText (txt,"%add:\s+","%add:-tab-") into txt
put replacetext (txt,"-tab-","tab") into txt
put replacetext (txt,"\.\.", "+...") into txt
put replacetext (txt,"?\.", "+..?") into txt
put replacetext (txt,"\./", "+//.") into txt
put replacetext (txt,"?\|", "+/?") into txt
put replacetext (txt,".\-", "+/." ) into txt
put replacetext (txt,"?\|-", "+/?") into txt
put replacetext (txt,"<p\|>", "(.)") into txt
put replacetext (txt,"<Lp\|>", "(.)") into txt
put replacetext (txt,"XXX", "xxx") into txt
put txt into field "Resultat"
end remplacement2

```

-- Effacement des balises superflues : erreurs enfantines, transcriptions phonétiques, doubles chevauchements et commentaires du transcripteur

```

on effacement
put the number of lines of field "Resultat" into nbl
repeat with x = 1 to nbl
  put line x of field "Resultat" into laligne
  repeat while matchText (laligne,"(*)<sic/>\s(*)",debut,fin) = true
    if matchText (laligne,"(*)<sic/>\s(*)",debut,fin) = true then
      put debut& \
        fin \
        into laligne
      put laligne into line x of field "Resultat"
    end if
  end repeat
end repeat

put the number of lines of field "Resultat" into nbl
repeat with x = 1 to nbl
  put line x of field "Resultat" into laligne
  repeat while matchText (laligne,"(*)<phon>\s\S+\s</phon>\s(*)",debut,fin) = true
    if matchText (laligne,"(*)<phon>\s\S+\s</phon>\s(*)",debut,fin) = true then
      put debut& \
        fin \
        into laligne
      put laligne into line x of field "Resultat"
    end if
  end repeat
end repeat

put the number of lines of field "Resultat" into nbl
repeat with x = 1 to nbl
  put line x of field "Resultat" into laligne
  repeat while matchText (laligne,"(*)<<(*)",debut,fin) = true
    if matchText (laligne,"(*)<<(*)",debut,fin) = true then
      put debut& \
        "<& \
        fin \
        into laligne

```

```

    put laligne into line x of field "Resultat"
  end if
end repeat
end repeat

put the number of lines of field "Resultat" into nbl
repeat with x = 1 to nbl
  put line x of field "Resultat" into laligne
  repeat while matchText (laligne,"<!--.+-->",debut,fin) = true
    if matchText (laligne,"<!--.+-->",debut,fin) = true then
      put "" into laligne
      put laligne into line x of field "Resultat"
    end if
  end repeat
end repeat

end effacement

```

-- Ordre d'application des différents codes
--

```
on mouseUp
```

```

  put field "Texte" into field "Resultat"
  traiterCasLiaison
  metBlanc
  traiterCasLangue
  metBlanc
  traitementCasVoix
  metBlanc
  traiterCasRepetition
  metBlanc
  traiterCasAlternative
  metBlanc
  traiterCasChevauchement
  metBlanc
  remplacements1
  metBlanc
  traitementAdresse
  metBlanc
  traitementMotsSpeciaux
  metBlanc
  remplacement2
  effacement

```

```
end mouseUp
```

Programme Perl de conversion du corpus structuré au format XML-CHAT
vers le format XML-TEI

Début du programme

```
use Encode qw(encode decode);
```

Préciser le nom du fichier à la place de XXX

```
open (A, "<:encoding(UTF-8)", "XXX.xml") || die "Impossible d'ouvrir le fichier\n" ;
while (<A>) {
    @corres = ($_ =~ /<.*\>/g) ;
    foreach $i (@corres) {
```

Conversion des balises contenant les informations sur les locuteurs

```
$_ =~ s/<u who=\")(MOT)(\")/$1\#$2\ -Baptiste$3/g ;
$_ =~ s/<u who=\")(FAT)(\")/$1\#$2\ -Baptiste$3/g ;
$_ =~ s/<u who=\")(CHI)(\")/$1\#$2\ -Baptiste$3/g ;
```

Conversion des balises la présence d'une pause

```
$_ =~ s/<pause symbolic-length="simple"\>/<pause\>/g ;
```

Gestion des transcriptions alternatives (1)

```
$_ =~ s/<ga type="alternative">(\w+) (\w+) (\w+) (\w+)
(\w+)(<\>/ga>)/$1<w>$2<\>/w> <w>$3<\>/w> <w>$4<\>/w> <w>$5<\>/w>
<w>$6<\>/w>$7/g ;
$_ =~ s/<ga type="alternative">(\w+) (\w+) (\w+)
(\w+)(<\>/ga>)/$1<w>$2<\>/w> <w>$3<\>/w> <w>$4<\>/w> <w>$5<\>/w>$6/g ;
$_ =~ s/<ga type="alternative">(\w+) (\w+) (\w+)(<\>/ga>)/$1<w>$2<\>/w>
<w>$3<\>/w> <w>$4<\>/w>$5/g ;
$_ =~ s/<ga type="alternative">(\w+) (\w+)(<\>/ga>)/$1<w>$2<\>/w>
<w>$3<\>/w>$4/g ;
$_ =~ s/<ga type="alternative">(\w+)(<\>/ga>)/$1<w>$2<\>/w>$3/g ;
```

Gestion des énoncés non transcrits car incompréhensibles

```
$_ =~ s/<w untranscribed="unintelligible">xxx</w>/<unclear extent="several words"/>/g ;
$_ =~ s/<w>xx</w>/<unclear extent="one word"/>/g ;
```

Conversion des balises marquant les élisions

```
$_ =~
s/(<w>)(\S+)<shortening>(\S+)</shortening>(\S+)<shortening>(\S+)</shortening>(<\w>)/$1<choice><orig>$2\'$4</orig><reg>$2$3$4$5</reg></choice>$6/g ;
$_ =~
s/(<w>)(\S+)<shortening>(\S+)</shortening>(\S+)(<\w>)/$1<choice><orig>$2\'$4</orig><reg>$2$3$4</reg></choice>$5/g ;
$_ =~
s/(<w>)<shortening>(\S+)</shortening>(\S+)(<\w>)/$1<choice><orig>\'$3</orig><reg>$2$3</reg></choice>$4/g ;
$_ =~
s/(<w>)(\S+)<shortening>(\S+)</shortening>(<\w>)/$1<choice><orig>$2\'</orig><reg>$2$3</reg></choice>$4/g ;
$_ =~
s/(<w>)(\S+)<shortening>(\S+)</shortening>(\S+)<shortening>(\S+)</shortening>(\S+)(<\w>)/$1<choice><orig>$2\'$4\'$6</orig><reg>$2$3$4$5$6</reg></choice>$7/g ;
```

Gestion des chevauchements d'énoncés

```
$_ =~ s/uID/xml:id/g ;
$_ =~ s/<overlap t\S+ \S+\"/>/g ;
$_ =~ s/<t t\S+\"/>/g ;
$_ =~ s/(unit=\"s\") (\>)/$1$2/g ;
$_ =~ s/(<u.*xml:id=\"(\S+)(\">)(.*)<media start=\"\S+\" end=\"\S+\" unit\S+(.*)(</u>)/$1$2$3 <anchor synch=\"\$2-start\"/>$4<anchor synch=\"\$2-end\"/>$5$6/g ;
$_ =~ s/(<u.*xml:id=\"(\S+)(\">)(.*)<internal-media start=\"\S+\" end=\"\S+\" unit\S+(.*)(</u>)/$1$2$3 <anchor synch=\"\$2-start\"/>$4<anchor synch=\"\$2-end\"/>$5$6/g ;
```

Gestion des annotations sur la liaison (1)

Plusieurs codes possibles en fonction des particularités de l'énoncé

```
$_ =~ s/<w>(\S+)</w> <freecode>Syntctx=(.) expecCons=(.) realCons=(.) obliOpt=(.)</freecode> <g><w>(\S+)</w><ga>(\S+</g>)/<seg> <w>$1</w> <fs type=\"liaison\"><f name=\"Word1\" fVal=\"\$1\"/><f name=\"Word2\" fVal=\"\$6\"/><f
```

```

name=\ "SynctacticContext\ " fVal=\ "$2\ \"/><f
name=\ "ExpectedConsonnant\ " fVal=\ "$3\ \"/><f
name=\ "ProducedConsonnant\ " fVal=\ "$4\ \"/><f
name=\ "ObligatoryOptional\ " fVal=\ "$5\ \"/><\ /fs> <g><w>$6<\ /w><ga
$7<\ /seg>/g ;

```

```

$ _ =~ s/<w>(\ S+)<\ /w> <freecode>Syntctx=(.) expecCons=(.)
realCons=(.) obliOpt=(.)<\ /freecode> <w>(\ S+)<\ /w>
<freecode>Syntctx=(.) expecCons=(.) realCons=(.) obliOpt=(.)<\ /freecode>
<w>(\ S+)<\ /w> <freecode>Syntctx=(.) expecCons=(.) realCons=(.)
obliOpt=(.)<\ /freecode> <w>(\ S+)<\ /w>/ <seg> <w>$1<\ /w> <fs
type=\ "liaison\ "><f name=\ "Word1\ " fVal=\ "$1\ \"/><f name=\ "Word2\ "
fVal=\ "$6\ \"/><f name=\ "SynctacticContext\ " fVal=\ "$2\ \"/><f
name=\ "ExpectedConsonnant\ " fVal=\ "$3\ \"/><f
name=\ "ProducedConsonnant\ " fVal=\ "$4\ \"/><f
name=\ "ObligatoryOptional\ " fVal=\ "$5\ \"/><\ /fs> <w>$6<\ /w> <fs
type=\ "liaison\ "><f name=\ "Word1\ " fVal=\ "$6\ \"/><f name=\ "Word2\ "
fVal=\ "$11\ \"/><f name=\ "SynctacticContext\ " fVal=\ "$7\ \"/><f
name=\ "ExpectedConsonnant\ " fVal=\ "$8\ \"/><f
name=\ "ProducedConsonnant\ " fVal=\ "$9\ \"/><f
name=\ "ObligatoryOptional\ " fVal=\ "$10\ \"/><\ /fs> <w>$11<\ /w> <fs
type=\ "liaison\ "><f name=\ "Word1\ " fVal=\ "$11\ \"/><f name=\ "Word2\ "
fVal=\ "$16\ \"/><f name=\ "SynctacticContext\ " fVal=\ "$12\ \"/><f
name=\ "ExpectedConsonnant\ " fVal=\ "$13\ \"/><f
name=\ "ProducedConsonnant\ " fVal=\ "$14\ \"/><f
name=\ "ObligatoryOptional\ " fVal=\ "$15\ \"/><\ /fs> <w>$16<\ /w>
<\ /seg>/g ;

```

```

$ _ =~ s/<w>(\ S+)<\ /w> <freecode>Syntctx=(.) expecCons=(.)
realCons=(.) obliOpt=(.)<\ /freecode> <w>(\ S+)<\ /w>
<freecode>Syntctx=(.) expecCons=(.) realCons=(.) obliOpt=(.)<\ /freecode>
<w>(\ S+)<\ /w>/<seg> <w>$1<\ /w> <fs type=\ "liaison\ "><f
name=\ "Word1\ " fVal=\ "$1\ \"/><f name=\ "Word2\ " fVal=\ "$6\ \"/><f
name=\ "SynctacticContext\ " fVal=\ "$2\ \"/><f
name=\ "ExpectedConsonnant\ " fVal=\ "$3\ \"/><f
name=\ "ProducedConsonnant\ " fVal=\ "$4\ \"/><f
name=\ "ObligatoryOptional\ " fVal=\ "$5\ \"/><\ /fs> <w>$6<\ /w> <fs
type=\ "liaison\ "><f name=\ "Word1\ " fVal=\ "$6\ \"/><f name=\ "Word2\ "
fVal=\ "$11\ \"/><f name=\ "SynctacticContext\ " fVal=\ "$7\ \"/><f
name=\ "ExpectedConsonnant\ " fVal=\ "$8\ \"/><f
name=\ "ProducedConsonnant\ " fVal=\ "$9\ \"/><f
name=\ "ObligatoryOptional\ " fVal=\ "$10\ \"/><\ /fs> <w>$11<\ /w>
<\ /seg>/g ;

```

```

$ _ =~ s/(<pause\ />) <freecode>Syntctx=(.) expecCons=(.)
realCons=(.) obliOpt=(.)<\ /freecode> <w>(\ S+)<\ /w>/<seg> $1 <fs
type=\ "liaison\ "><f name=\ "Word1\ " fVal=\ "PAUSE\ \"/><f
name=\ "Word2\ " fVal=\ "$6\ \"/><f name=\ "SynctacticContext\ "
fVal=\ "$2\ \"/><f name=\ "ExpectedConsonnant\ " fVal=\ "$3\ \"/><f
name=\ "ProducedConsonnant\ " fVal=\ "$4\ \"/><f
name=\ "ObligatoryOptional\ " fVal=\ "$5\ \"/><\ /fs> $6 <\ /seg>/g ;

```

```

    $_ =~ s/<w>(\S+)</w> <freecode>Syntctx=(.) expecCons=(.)
realCons=(.) obliOpt=(.)</freecode> <w>(\S+)</w>/<seg>
<w>$1</w> <fs type="liaison"><f name="Word1" fVal="$1"/><f
name="Word2" fVal="$6"/><f name="SynctacticContext"
fVal="$2"/><f name="ExpectedConsonnant" fVal="$3"/><f
name="ProducedConsonnant" fVal="$4"/><f
name="ObligatoryOptional" fVal="$5"/></fs> <w>$6</w>
</seg>/g ;
    $_ =~ s/(start"/>.* )<freecode>Syntctx=(.) expecCons=(.)
realCons=(.) obliOpt=(.)</freecode> <w>(\S+)</w>/$1<seg><fs
type="liaison"><f name="Word1" fVal="O"/><f name="Word2"
fVal="$6"/><f name="SynctacticContext" fVal="$2"/><f
name="ExpectedConsonnant" fVal="$3"/><f
name="ProducedConsonnant" fVal="$4"/><f
name="ObligatoryOptional" fVal="$5"/></fs> <w>$6</w>
</seg>/g ;

```

Gestion des transcriptions alternatives (2)

```

    $_ =~ s/<(u.*xml:id=")(\S+)(\".*)<g><(w>\S+</w><ga
type="alternative">(\S+)</ga></g>(.)/$1$2$3<seg exclude="\#2b"
xml:id="\#2a">$4</seg><seg exclude="\#2a" xml:id="\#2b">$5</seg>$6/g
;
    $_ =~ s/<(u.*xml:id=")(\S+)(\".*)<g><(w>\S+</w> <w>\S+</w><ga
type="alternative">(\S+)</ga></g>(.)/$1$2$3<seg exclude="\#2b"
xml:id="\#2a">$4</seg><seg exclude="\#2a" xml:id="\#2b">$5</seg>$6/g
;
    $_ =~ s/<(u.*xml:id=")(\S+)(\".*)<g><(w>\S+</w> <w>\S+</w>
<w>\S+</w><ga type="alternative">(.)</ga></g>(.)/$1$2$3<seg
exclude="\#2b" xml:id="\#2a">$4</seg><seg exclude="\#2a"
xml:id="\#2b">$5</seg>$6/g ;
    $_ =~ s/<(u.*xml:id=")(\S+)(\".*)<g>(.)<ga
type="alternative">(.)</ga></g>(.)/$1$2$3<seg exclude="\#2b"
xml:id="\#2a">$4</seg><seg exclude="\#2a" xml:id="\#2b">$5</seg>$6/g
;

```

Conversion des balises contenant les informations sur les modes de production des énoncés

```

    $_ =~ s/<e><action/><ga type="paralinguistics">rit</ga></e>/<vocal
type="\laughs"/>/g ;
    $_ =~ s/<e><action/><ga type="paralinguistics">pleure</ga></e>/<vocal
type="\cries"/>/g ;
    $_ =~ s/<e><action/><ga type="paralinguistics">crie</ga></e>/<vocal
type="\screams"/>/g ;
    $_ =~ s/<e><action/><ga type="paralinguistics">g.mit</ga></e>/<vocal
type="\moans"/>/g ;
    $_ =~ s/<e><action/><ga type="paralinguistics">tousse</ga></e>/<vocal
type="\coughs"/>/g ;

```

```

    $_ =~ s/<e><action\\/><ga type="paralinguistics">soupire<\\ga><\\e>/<vocal
type\\="sighs\\"/>/g ;
    $_ =~ s/<e><action\\/><ga type="paralinguistics">chante<\\ga><\\e>/<vocal
type\\="sings\\"/>/g ;
    $_ =~ s/<e><action\\/><ga type="paralinguistics">baille<\\ga><\\e>/<vocal
type\\="yawns\\"/>/g ;

```

```

    $_ =~ s/<g>(.*<ga type="paralinguistics">crie<\\ga><\\g>/<shift
new="screaming\\"/>$1<shift\\/>/g ;
    $_ =~ s/<g>(.*<ga type="paralinguistics">lit<\\ga><\\g>/<shift
new="reading\\"/>$1<shift\\/>/g ;
    $_ =~ s/<g>(.*<ga type="paralinguistics">chante<\\ga><\\g>/<shift
new="singing\\"/>$1<shift\\/>/g ;
    $_ =~ s/<g>(.*<ga type="paralinguistics">chuchote<\\ga><\\g>/<shift
new="whispering\\"/>$1<shift\\/>/g ;
    $_ =~ s/<g>(.*<ga type="paralinguistics">rit<\\ga><\\g>/<shift
new="laughing\\"/>$1<shift\\/>/g ;
    $_ =~ s/<g>(.*<ga type="paralinguistics">pleure<\\ga><\\g>/<shift
new="crying\\"/>$1<shift\\/>/g ;
    $_ =~ s/<g>(.*<ga type="paralinguistics">g.mit<\\ga><\\g>/<shift
new="moaning\\"/>$1<shift\\/>/g ;
    $_ =~ s/<g>(.*<ga type="paralinguistics">baille<\\ga><\\g>/<shift
new="yawning\\"/>$1<shift\\/>/g ;
    $_ =~ s/<g>(.*<ga type="paralinguistics">tousse<\\ga><\\g>/<shift
new="coughing\\"/>$1<shift\\/>/g ;
    $_ =~ s/<g>(.*<ga type="paralinguistics">soupire<\\ga><\\g>/<shift
new="sighing\\"/>$1<shift\\/>/g ;

```

Conversion des balises contenant les informations sur l'adresse du discours

```

    $_ =~ s/<u .*><a type="addressee">.(\\S+)<\\a>( <\\u>)/$1<fs
type="addressee"><f name="target" fVal="$2\\"/><\\fs>$3/g ;

```

Rattrapage des erreurs comme les doubles espacements par exemple

```

    $_ =~ s/>>>/>/g ;
    $_ =~ s/>></</g ;
    $_ =~ s/<g>//g ;
    $_ =~ s/<\\g>//g ;
    $_ =~ s/<t ty.+\\t>//g ;
    $_ =~ s/<linker type=\\S+ \\S+\\"/>//g ;
    $_ =~ s/ / /g ;

```

Conversion des balises contenant les informations sur les formes spéciales

```

    $_ =~ s/<w formType="(\\S+)">(\\S+<\\w>)/<w subtype="$1\\">$2/g ;

```

Gestion des annotations sur la liaison (1)

```
$_ =~ s/(<f name="Word.\"  
fVal="\")<choice><orig>\S+</orig><reg>(\S+)</reg></choice>("\/>)/$1$2$3/g ;  
$_ =~  
s/(<w><shortening>(\S+)</shortening>(</w>))/ $1<choice><orig\/><reg>$2</reg>  
</choice>$3/g ;
```

Conversion des balises contenant les informations sur la langue des énoncés produits

```
$_ =~ s/<w><langs><single>eng</single></langs>(\S+)</w>/<w  
xml:lang="en-GB">$1</w>/g ;
```

```
$ligne = encode ('UTF-8', "$_") ;  
print "$ligne" ;  
}  
}
```

```
close(A) ;
```


Programmes d'extraction des informations sur la liaison

Il y a cinq programmes à appliquer afin d'obtenir un tableau contenant l'ensemble des informations sur la liaison que nous avons annotées. Ceux-ci peuvent être appelés dans un fichier de commandes Batch en suivant l'ordre et la nomenclature des fichiers suivants afin de les déclencher successivement :

```
perl corpus.pl > corpus+.xml
perl retouches1.pl > corpus++.xml
perl retouches2.pl > corpus+++.xml
perl conversion_tableau.pl > tableau.xml
perl m1.pl > tableau.txt
```

corpus.pl

```
use Encode qw(encode decode);

# Remplacer XXX par le nom du fichier XML à analyser

open (A, "<:encoding(UTF-8)", "XXX.xml") || die "Impossible d'ouvrir le fichier xml\n" ;

while (<A> ) {
    $_ =~ s/<ens.>//g ;
    $_ =~ s/<\/ens.>//g ;
    $_ =~ s/(\[^\^]+\^)\s*(<...\/>)\s*([\^s]*)/<m1>$1<\/m1> $2 <m2>$3<\/m2>/g ;
    $_ =~ s/(c'est)\s*(<...\/>)\s*([\^s]*)/<m1>$1<\/m1> $2 <m2>$3<\/m2>/g ;
    $_ =~ s/([\^s'<>]+\^)\s*(<...\/>)\s*([\^s]*)/<m1>$1<\/m1> $2 <m2>$3<\/m2>/g ;
    $_ =~ s/(<m2>[\^>]+\^<\/m2>)\s*(<...\/>)\s*([\^s]*)/$1 $2 <m2>$3<\/m2>/g ;
    $_ =~ s/([\^s]*)\s*(<ens>\s*<...\/>)\s*([\^s]*)/<m1>$1<\/m1> $2 <m2>$3<\/m2>/g ;
    $_ =~ s/([\^s]*)\s*(<\/ens>\s*<...\/>)\s*([\^s]*)/<m1>$1<\/m1> $2 <m2>$3<\/m2>/g ;
    $_ =~ s/([\^s]*)\s*(<phon>\s*[\^s]*\s*<\/phon>\s*<...\/>)\s*([\^s]*)/<m1>$1<\/m1> $2 <m2>$3<\/m2>/g ;
    $_ =~ s/(<p\/>)\s*(<...\/>)\s*([\^s]*)/<m1>$1<\/m1> $2 <m2>$3<\/m2>/g ;
    $ligne = encode('UTF-8', "$_") ;
    print "$ligne" ;
}

close(A) ;
```

retouches1.pl

```
use Encode qw(encode decode);

open (A, "<:encoding(UTF-8)", "corpus+.xml") || die "Impossible d'ouvrir le fichier\n" ;
```

```

while (<A>) {
$_ =~ s/<m2><\ens><\m2> \s*([\s]*)/<m2>$1<\m2>/g ;
$_ =~ s/<\ens>//g ;
$_ =~ s/<ens>//g ;
$_ =~ s/<m2><\m2> \s*([\s]*)/<m2>$1<\m2>/g ;
$_ =~ s/<m2><m1>([\s]*)<\m1><\m2>/<m2>$1<\m2>/g ;
$_ =~ s/<m1><m2>([\s]*)<\m2><\m1>/<m2>$1<\m2>/g ;

    $ligne = encode('UTF-8','$_') ;
    print "$ligne" ;

}

close(A) ;

```

retouches2.pl

```
use Encode qw(encode decode);
```

```
open (A, "<:encoding(UTF-8)", "corpus++.xml") || die "Impossible d'ouvrir le fichier\n" ;
```

```

while (<A>) {
$_ =~ s/<m1>tit<\m1>/<m1>petit<\m1>/g ;
$_ =~ s/<m1>tits<\m1>/<m1>petits<\m1>/g ;
$_ =~ s/<m1>tites<\m1>/<m1>petites<\m1>/g ;
$_ =~ s/<m2>tit<\m2>/<m2>petit<\m2>/g ;
$_ =~ s/<m2>tits<\m2>/<m1>petits<\m2>/g ;
$_ =~ s/<m2>tites<\m2>/<m2>petites<\m2>/g ;

    $ligne = encode('UTF-8','$_') ;
    print "$ligne" ;

}

close(A) ;

```

conversion_tableau.pl

```
use Encode qw(encode decode);
```

```
print "NUM;LOC;MOT1;MOT2;CONTEXTE;CL;CA;J/F;LF/LO;ADRESSE;\n" ;
```

```
open (A, "<:encoding(UTF-8)", "corpus+++.xml") || die "Impossible d'ouvrir le fichier\n" ;
```

```

while (<A>) {
    @corres = ($_ =~ /<m1>[\^<>]+<\m1>\s*<....\>\s*<m2>[\^<>]+<\m2>/g) ;
    @corres2 = ($_ =~ /<m2>[\^<>]+<\m2>\s*<....\>\s*<m2>[\^<>]+<\m2>/g) ;
    if ($_ =~ /<m2>[\^<>]+<\m2>\s*<....\>\s*(<m2>[\^<>]+<\m2>\s*<....\>\s*<m2>[\^<>]+<\m2>)/g) {
        @corres3 = $1 ;
    } else {
        @corres3 = () ;
    }
}

```

```

    @corres6 = ($_ =~
/<m1>[^\<>]+<vm1>\s*<phon>\s*[\^s]*\s*<vphon>\s*<...v>\s*<m2>[^\<>]+<vm2>/g) ;
    push(@corres,@corres2,@corres3,@corres4,@corres5,@corres6) ;
    foreach $i (@corres) {
        $compte++;
        @loc = ($_ =~ /<(L[123])>/) ;
        @disc = ($_ =~ /<(A.)\>/g) ;

        $i =~ /([\^s]*)\s*<(.)(<.>(<.>(<.>)\>)\s*([\^s]*)/ ;
        $i =~ /([\^s]*)\s*<phon>\s*[\^s]*\s*<vphon>\s*<(.)(<.>(<.>(<.>)\>)\s*([\^s]*)/ ;
        $ca = $3 ;
        $cr = $4 ;
        $ob = $5 ;
        if ($cr eq $ca && $ob eq "0") {
            $eval = "J liaison facultative rÈalisÈe juste" ;
        }
        elsif ($cr eq $ca && $ob eq "1") {
            $eval = "J liaison obligatoire rÈalisÈe juste" ;
        }
        elsif ($cr eq "o" && $ob eq "0") {
            $eval = "J liaison facultative non rÈalisÈe" ;
        }
        elsif ($cr eq "o" && $ob eq "1") {
            $eval = "F liaison obligatoire non rÈalisÈe" ;
        }
        elsif ($cr ne "o" && $cr ne $ca && $ob eq "0") {
            $eval = "F liaison facultative rÈalisÈe fausse" ;
        }
        elsif ($cr ne "o" && $cr ne $ca && $ob eq "1") {
            $eval = "F liaison obligatoire rÈalisÈe fausse" ;
        }
        elsif ($ca eq "o" && $cr ne "o") {
            $eval = "F liaison abusive" ;
        }
        else {
            $eval = "F" ;
        }
        $sortie = encode("UTF-8", "$compte;$loc[0];$1;$6;$2;$4;$3;$eval;$5;$disc[0];\n") ;
        print "$sortie" ;
    }
}

close(A) ;

```

corpus.pl

```
use Encode qw(encode decode);
```

```
open (A, "<:encoding(UTF-8)", "tableau.xml") || die "Impossible d'ouvrir le fichier\n" ;
```

```
while (<A>) {
    $_ =~ s/<m[12]>(\D+)<vm[12]>/$1/g ;
```

```
print "$_" ;
```

```
}
```

```
close(A);
```



```

$_ =~ s/<ens2>//g ;
$_ =~ s/<\/ens2>//g ;
$_ =~ s/<r>//g ;
$_ =~ s/<\/r>//g ;
$_ =~ s/<rrr>//g ;
$_ =~ s/<\/rrr>//g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;
$_ =~ s/^(q*[jtmlscnud](e))\s/ <0>$1<\/0> /g ;

```

```

    $ligne = encode('UTF-8',"$_");
    print "$ligne" ;
}

```

```
close(A) ;
```

decoupe_elis.pl


```

$_ =~ s/(.*)(<0>)(\S+)(</0>)(.*)(<0>)(\S+)(</0>)(.*)(<0>)(\S+)(</0>)(.*)
(<0>)(\S+)(</0>)(.*)/$1$2$3$4$5$7$9$11$13$15$17\n$1$3$5$6$7$8$9$11$13$15$17\n$1$3$5$7$9$10$11$
12$13$15$17\n$1$3$5$7$9$11$13$14$15$16$17/g ;
$_ =~ s/(.*)(<0>)(\S+)(</0>)(.*)(<0>)(\S+)(</0>)(
.*)(<0>)(\S+)(</0>)(.*)/$1$2$3$4$5$7$9$11$13\n$1$3$5$6$7$8$9$11$13\n$1$3$5$7$9$10$11$12$13/g ;
$_ =~ s/(.*)(<0>)(\S+)(</0>)(.*)(<0>)(\S+)(</0>)(.*)/$1$2$3$4$5$7$9\n$1$3$5$6$7$8$9/g ;

$ligne = encode('UTF-8',"$_") ;
print "$ligne" ;
}

close(A) ;

```

tab-elis.pl

```

use Encode qw(encode decode);

print "NUM;LOC;CtxG;M1;CtxD;ELIS;ADRESSE;\n" ;

open (A, "<:encoding(UTF-8)", "elis-decoupe.xml") || die "Impossible d'ouvrir le fichier\n" ;

while (<A>) {
    @corres = ($_ =~ /.*<0>\S+</0>.*</g>) ;

    foreach $i (@corres) {
        $compte++ ;
        @loc = ($_ =~ /^<(L[123])>/) ;
        @disc = ($_ =~ /<(A.)\>/g) ;
        $i =~ /(\S+\s*\S*\s*\S*) <(0)>(\S+)</0> (\S+\s*\S*\s*\S*)/g ;
        $m1 = $3 ;
        $elis = $2 ;
        $ctxG = $1 ;
        $ctxD = $4 ;
        $sortie = encode('UTF-8',"$compte;$loc[0];$ctxG;$m1;$ctxD;$elis;$disc[0];\n") ;
        print "$sortie" ;
    }
}

close(A) ;

```

extraction-motGD.pl

```

use Encode qw(encode decode);

print "NUM;LOC;MotG;M1;MotD;ELIS;ADRESSE;\n" ;

open (A, "<:encoding(UTF-8)", "elis-decoupe.xml") || die "Impossible d'ouvrir le fichier\n" ;

while (<A>) {
    @corres = ($_ =~ /.*<0>\S+</0>.*</g>) ;

    foreach $i (@corres) {
        $compte++ ;
        @loc = ($_ =~ /^<(L[123])>/) ;
        @disc = ($_ =~ /<(A.)\>/g) ;

```

```
$i =~ /(\S+)<(0)>(\S+)<\0> (\S+)/g ;
$m1 = $3 ;
$e1is = $2 ;
$MotG = $1;
$MotD = $4;
$sortie = encode("UTF-8","$compte;$loc[0];$MotG;$m1;$MotD;$e1is;$disc[0];\n") ;
print "$sortie" ;
}
}
close(A) ;
```



```

$ _ =~ s/(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)
)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)
)(<1>(\S+)(<\/1>)(.*)/$1$2$3$4$5$7$9$11$13$15$17$19$21$23$25$27$29\n$1$3$5$6$7$8$9$11$13$15$17
$19$21$23$25$27$29\n$1$3$5$7$9$10$11$12$13$15$17$19$21$23$25$27$29\n$1$3$5$7$9$11$13$14$15$
16$17$19$21$23$25$27$29\n$1$3$5$7$9$11$13$15$17$18$19$20$21$23$25$27$29\n$1$3$5$7$9$11$13$1
5$17$19$21$22$23$24$25$27$29\n$1$3$5$7$9$11$13$15$17$19$21$23$25$26$27$28$29/g ;
$ _ =~ s/(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)
)(<1>(\S+)(<\/1>)(.*)
)(<1>(\S+)(<\/1>)(.*)/$1$2$3$4$5$7$9$11$13$15$17$19$21$23$25\n$1$3$5$6$7$8$9$11$13$15$17$19$21
$23$25\n$1$3$5$7$9$10$11$12$13$15$17$19$21$23$25\n$1$3$5$7$9$11$13$14$15$16$17$19$21$23$25\
n$1$3$5$7$9$11$13$15$17$18$19$20$21$23$25\n$1$3$5$7$9$11$13$15$17$19$21$22$23$24$25/g ;
$ _ =~ s/(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)
)(<1>(\S+)(<\/1>)(.*)/$1$2$3$4$5$7$9$11$13$15$17$19$21\n$1$3$5$6$7$8$9$11$13$15$17$19$21\n$1$3$
5$7$9$10$11$12$13$15$17$19$21\n$1$3$5$7$9$11$13$14$15$16$17$19$21\n$1$3$5$7$9$11$13$15$17$1
8$19$20$21/g ;
$ _ =~ s/(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)
)(<1>(\S+)(<\/1>)(.*)/$1$2$3$4$5$7$9$11$13$15$17\n$1$3$5$6$7$8$9$11$13$15$17\n$1$3$5$7$9$10$11$
12$13$15$17\n$1$3$5$7$9$11$13$14$15$16$17/g ;
$ _ =~ s/(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(
.*)(<1>(\S+)(<\/1>)(.*)/$1$2$3$4$5$7$9$11$13\n$1$3$5$6$7$8$9$11$13\n$1$3$5$7$9$10$11$12$13/g ;
$ _ =~ s/(.*)(<1>(\S+)(<\/1>)(.*)(<1>(\S+)(<\/1>)(.*)/$1$2$3$4$5$7$9\n$1$3$5$6$7$8$9/g ;

$ligne = encode('UTF-8',"$_") ;
print "$ligne" ;
}

close(A) ;

```

tab-maintien.pl

```

use Encode qw(encode decode);

print "NUM;LOC;CtxG;M1;CtxD;ELIS;ADRESSE;\n" ;

open (A, "<:encoding(UTF-8)", "maintien-decoupe.xml") || die "Impossible d'ouvrir le fichier\n" ;

while (<A> ) {
    @corres = ($ _ =~ /.*<1>\S+<\/1>.*\/g) ;

    foreach $i (@corres) {
        $compte++ ;
        @loc = ($ _ =~ /^<(L[123])>\/) ;
        @disc = ($ _ =~ /<(A.)\>\/g) ;
        # $i =~ /(\S+)<(1)>(\S+)<\/1> (\S+)/ ;
        $i =~ /(\S+\s*\S*\s*\S*) <(1)>(\S+)<\/1> (\S+\s*\S*\s*\S*)\/g ;
        $m1 = $3 ;
        $elis = $2 ;
        $ctxG = $1 ;
        $ctxD = $4 ;
        $sortie = encode('UTF-8',"$compte;$loc[0];$ctxG;$m1;$ctxD;$elis;$disc[0];\n") ;
        print "$sortie" ;
    }
}

close(A) ;

```

extraction-motGD.pl

```

use Encode qw(encode decode);

print "NUM;LOC;MotG;M1;MotD;ELIS;ADRESSE;\n" ;

open (A, "<:encoding(UTF-8)", "maintien-decoupe.xml") || die "Impossible d'ouvrir le fichier\n" ;

while (<A> ) {
    @corres = ($_ =~ /.*<1>\S+<\/1>.*\/g) ;

    foreach $i (@corres) {
        $compte++ ;
        @loc = ($_ =~ /^<(L[123])>\/) ;
        @disc = ($_ =~ /<(A.)\/>\/g) ;
        $i =~ /(\S+) <(1)>(\S+)<\/1> (\S+)/g ;
        $m1 = $3 ;
        $elis = $2 ;
        $MotG = $1 ;
        $MotD = $4 ;
        $sortie = encode("UTF-8", "$compte;$loc[0];$MotG;$m1;$MotD;$elis;$disc[0];\n") ;
        print "$sortie" ;
    }
}

close(A) ;

```