

Une introduction au système CHILDES en français

CHILDES : Child Language Data Exchange System (Système d'échange de données sur le langage de l'enfant)

Site pour l'Europe (Université d'Anvers) : <http://atila-www.uia.ac.be/childes/>

“ Bulletin board ” électronique : info-childes@mail.talkbank.org

Pour souscrire : <mailto:info-childes-on@mail.talkbank.org>

Référence (2 volumes + un CD-ROM) :

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk, Third edition. Volume I: Transcription format and programs*. Mahwah, NJ: Lawrence Erlbaum Associates. ISBN 0-8058-2995-4

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk, Third edition. Volume II: The database*. Mahwah, NJ: Lawrence Erlbaum Associates. ISBN 0-8058-3572-5

Le système CHILDES se compose de trois outils :

- **(1) CHAT** : Codes for the **H**uman **A**nalysis of **T**ranscripts (principes systématiques de transcription et de codage) ; (voir infra p. 2, exemple p. 5, indications pour la transcription du français pp. 6-11)
- **(2) CLAN** : Computerized **L**anguage **A**nalysis (“ éditeur ” (traitement de texte) du CHILDES (sous Windows ou Macintosh), automatisant un grand nombre de procédures de transcription ou de codage et permettant la liaison des fichiers transcrits avec des fichiers son ou vidéo numérisés ; programmes de contrôle et d'analyse automatisée des corpus ainsi obtenus) ; (voir infra p. 3, exemples p. 4 et liste p. 12)
- **(3) Database** (banque de données) : corpus enregistrés et transcrits de productions (la plupart du temps spontanées) d'enfants, et concernant une trentaine de langues ; en plus des données documentant le développement normal de l'enfant dans une langue particulière, sont aussi inclus des corpus provenant d'enfants ayant des problèmes de langage (Down et Williams syndromes, autisme, SLI, troubles de l'articulation, etc.), d'adultes aphasiques, de bilingues, et d'apprenants de langue seconde. Outre les corpus en production, le CHILDES dispose aussi de données provenant de rapports parentaux (CDI : Communicative Development Inventory) ;
- **(+ Chibib)** : à ceci il faut ajouter une banque de données bibliographiques concernant l'acquisition du langage, comportant plus de 30 000 références, consultable ou téléchargeable sur le site du CHILDES.

Les manuels, les programmes, la banque de données, la bibliographie, ainsi que d'autres outils (par exemple polices de caractères IPA (IPAPhon, cf. p. 13 de ce document) ou facilités destinées à intégrer les aspects audiovisuels aux transcriptions) sont régulièrement mis à jour et peuvent être consultés ou téléchargés sans conditions sur le site du CHILDES. Cependant, l'utilisation de ces données doit se conformer aux principes du partage de données (*data sharing*) : confidentialité en ce qui concerne les personnes enregistrées (pseudonymes) et obligation de citer les chercheurs qui ont recueilli et accepté de partager leurs données lorsqu'elles sont utilisées dans des publications.

CHAT : Codes for the **H**uman Analysis of Transcripts, (Codes pour l'Analyse Humaine des Transcriptions ; cf. brève présentation supra p. 1 et indications de transcription p. 6 sqq.)

Trois principes :

- **(1)** Chaque énoncé est transcrit en tant qu'une entrée séparée dans le système. Dans le cas où un locuteur produit plusieurs énoncés successifs, ils doivent chacun correspondre à une “ **ligne principale** ” (*main line*) différente. Ceci est important pour pouvoir faire correspondre des informations additionnelles ou des codages à la ligne principale. Chaque ligne principale commence par un astérisque (*) suivi d'un trigramme en majuscules indiquant le locuteur, de deux points, d'un tabulateur et enfin de l'énoncé, terminé obligatoirement par un signe de ponctuation. Aucune ponctuation ne doit figurer à l'intérieur de l'énoncé ; les points de suspension, omniprésents dans le français écrit, sont bannis.

Exemple :

***MAY: t' as fait un gros dodo ?**

- **(2)** Les informations supplémentaires, les commentaires éventuels ainsi que les codages, sont entrés dans des “ lignes secondaires ” ou “ **lignes dépendantes** ” (*dependent tiers* : mot à mot "étages dépendants"), et doivent suivre immédiatement la ligne principale dont elles dépendent. Ces lignes dépendantes commencent par le symbole %, suivi d'un trigramme en minuscules, de deux points, d'un tabulateur et du contenu de la ligne. De nombreux types de *dependent tiers* sont prédéfinis dans le CHILDES, par exemple %act indiquera les actions du locuteur. Le CHILDES propose aussi des systèmes de codage systématique pour la phonologie (%pho), les actes de parole (%spa), la morphologie (%mor) et la syntaxe (%syn). Les utilisateurs du système peuvent en outre créer les systèmes de codages qu'ils jugent les plus appropriés à leur recherche.

- **(3)** Sur la ligne principale, (*) le but est d'entrer des formes qui correspondent le plus directement possible aux mots produits par le locuteur, en les séparant les uns des autres par un espace. Bien entendu, les formes produites par les enfants diffèrent sur de nombreux points des formes standard du langage, et il existe une grande variété de techniques dans le CHILDES pour noter ces divergences, tout en maintenant les possibilités de traitement informatique. Bien qu'il soit possible de coder directement sur la ligne principale certaines caractéristiques (phonétiques, morphologiques, etc.), il est préférable de le faire sur des lignes dépendantes spécifiques (en l'occurrence %pho et %mor). En effet, la ligne principale n'a pas vocation à tout coder, sinon elle devient rapidement illisible et difficile à analyser.

Outre les lignes principales et les lignes dépendantes, la troisième composante fondamentale d'un fichier CHAT est constituée par les “ **Headers** ”, ainsi appelés parce que certains d'entre eux doivent obligatoirement figurer en tête du fichier, et contiennent des informations qui restent constantes et valent pour tout le fichier. Ces *headers* ou *En-Têtes* commencent tous par un arrobas (@), et contiennent des informations sur la séance, sa date, l'endroit où elle se passe, les participants, etc. Tous les fichiers doivent commencer par une ligne @Begin et finir par une ligne @End. La deuxième ligne du fichier doit énumérer les participants ; ces trois headers sont obligatoires, et leur absence empêche les programmes de fonctionner. D'autres sont facultatifs, mais apportent des informations importantes, comme l'âge de l'enfant étudié (*Target_Child*).

Les fichiers ainsi transcrits selon le format CHAT doivent avoir l'extension *.cha*. Ils correspondent au format *.txt*, (texte seul) et peuvent donc être repris, mis en page, paginés, imprimés, etc., sous Word ; le programme MAKEDATA (cf. p. 12) permet d'effectuer les diverses conversions entre Windows, Macintosh, DOS et Unix. Exemple de nom de fichier CHAT : *antoine060.cha*.

Pour plus de détails, voir l'exemple de fichier p. 5 de ce texte. Pour des indications spécifiques à la transcription du français, voir pp. 6-11.

CLAN : Computerized Language Analysis (Analyse Informatique du Langage ; cf. brève présentation supra p. 1 ; présentation de l'éditeur du CHILDES p. 6 ; liste des programmes p. 12)

Les programmes **CLAN** sont construits spécifiquement pour analyser les données transcrites dans le format du CHILDES. Ces programmes ont été écrits par Leonid Spektor à Carnegie Mellon University (Pittsburgh, Pennsylvania, USA) et permettent d'effectuer de nombreuses analyses automatiques sur un ou plusieurs fichiers.

(exemples montrant les propriétés de la fenêtre "Commandes" p. 6 et Manuel Clan, pp. 3-14)

CLAN comporte un *editor* (traitement de texte) spécialement conçu pour travailler sur des fichiers codés en **CHAT** (Manuel **CLAN** pp. 15-31), et dans lequel de nombreuses procédures de transcription et de codage sont automatisées. Cet "éditeur" peut fonctionner selon plusieurs modes adaptés à différentes fonctions (**CHAT** mode pour la transcription, Coder mode pour le codage, Sonic mode, Video mode pour relier les transcriptions à des fichiers audio ou vidéo numérisés, etc.)

Lors de la transcription, il est important de vérifier souvent que l'on n'a pas commis d'erreurs. Ceci se fait automatiquement en lançant le programme **CHECK** sans sortir du fichier sur lequel on travaille (en cliquant sur "Mode" puis sur "Check open file"); **CHECK** indique alors les endroits qui comportent des erreurs et le type d'erreur commis.

Une fois le fichier transcrit, de nombreux programmes permettent d'obtenir des analyses ou des informations (cf. liste p. 12), de les conserver dans des fichiers séparés, de les imprimer, d'effectuer de nouvelles recherches ou de nouvelles analyses sur les résultats, etc. Il suffit d'entrer le nom du programme dans la fenêtre "commandes", de spécifier les options choisies si nécessaire et d'entrer le nom du ou des fichiers à analyser. Les programmes les plus utilisés dans notre travail sont les suivants :

- **CHECK** déjà présenté permet de vérifier si les fichiers ne comportent pas d'erreurs qui pourraient gêner le traitement (absence de ponctuation, parenthèses manquantes, caractères ou lignes non déclarées, etc.) ; en déclenchant **CHECK** à partir de la fenêtre "Commandes", on peut vérifier un grand nombre de fichiers à la fois. **CHECK** n'est pas un correcteur orthographique : la personne qui transcrit est libre de choisir son orthographe en ce qui concerne les mots produits par les locuteurs, mais l'expérience montre qu'une trop grande variabilité orthographique pour un mot donné engendre des difficultés de lecture et d'analyse.

- **COMBO** cherche les occurrences ou les co-occurrences de certains mots ou parties de mot ; selon les options, il fournit les lignes dans lesquelles ils apparaissent, avec éventuellement un contexte plus ou moins riche selon les options choisies.

- **DATES** calcule l'âge d'un sujet à partir de sa date de naissance et de la date de la session.

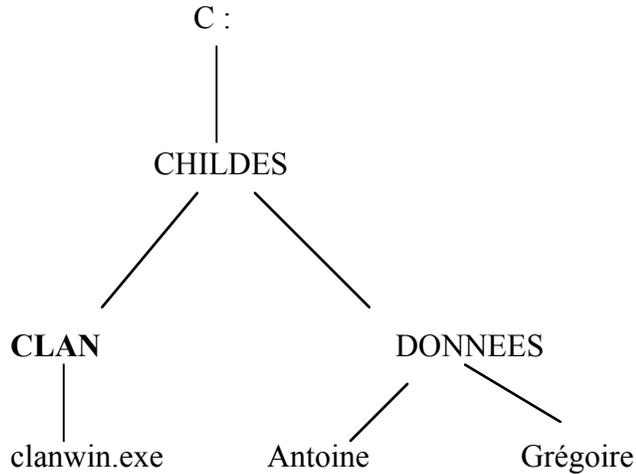
- **FREQ** calcule la fréquence d'un, de plusieurs, ou de tous les mots dans un fichier ou une série de fichiers. Il fournit d'autres informations, comme la fréquence des "types", des "tokens" et le rapport Type/Token, indice de diversité lexicale. **FREQ** permet aussi d'effectuer des recherches ou des statistiques sur des codes que l'on a entrés dans des *dependent tiers*. C'est un programme important, et pas seulement par les résultats quantitatifs qu'il fournit : le fait d'examiner la liste des mots avec leurs fréquences respectives permet, par exemple, de se rendre compte des variantes orthographiques non voulues par le transcripteur.

- **KWAL** retrouve et imprime les énoncés contenant certains mots auxquels on s'intéresse.

- **MLU** calcule la longueur moyenne des énoncés (Mean Length of Utterance), en mots ou en morphèmes, pour un locuteur donné, sur un ou plusieurs fichiers, ainsi que son écart-type. Il fournit d'autres renseignements comme le nombre total d'énoncés, le nombre total de mots (vs de morphèmes) sur la portion du corpus considérée.

- (voir p. 12 de ce document le tableau général et la description sommaire des programmes **CLAN**)

Exemple de configuration des répertoires sous Windows :



Bien entendu, les programmes CLAN peuvent travailler sur tous les répertoires ; la configuration proposée n'est donc qu'une solution que nous avons trouvée particulièrement fonctionnelle.

Exemple de commande et résultat :

freq +t*ANT antoine060.cha

Cette ligne de commandes comporte trois éléments séparés par des espaces. Le premier terme donne le nom du programme utilisé (**FREQ**). Le second terme spécifie que l'analyse doit porter uniquement sur la ligne principale correspondant au sujet Antoine (***ANT**). Le dernier terme indique au programme quel fichier il doit analyser (**antoine060.cha**, donné comme exemple p. 5). Voici le résultat de cette analyse :

```

> freq +t*ANT antoine060.cha
freq +t*ANT antoine060.cha
Tue Mar 27 14:33:43 2001
freq (11-NOV-99) is conducting analyses on:
  ONLY speaker main tiers matching: *ANT;
*****
From file <c:\CHILDES\Données\Antoine\ENCOURS\antoine060.cha>
  2 antoine
  2 c'
  1 clara
  2 dis
  1 dit
  1 elle
  2 est
  2 je
  1 là+haut
  1 maman
  2 non
 13 oui
  1 petit
  2 toujours
  1 trop
  2 à
-----
  16 Total number of different word types used
  36 Total number of words (tokens)
0.444 Type/Token ratio

```

Exemple de fichier au format CHAT :

@Begin
 @Participants: ANT Antoine Target_Child, CLA Clara Child, MAY Maya
 Investigator, JOS Josiane Adult
 @Filename: antoine060.cha
 @Date: 02-NOV-1995
 @Birth of ANT: 3-MAY-1993
 @Age of ANT: 2;5.30
 @Tape Location: cassette vidéo 8mm N° 8 (original) ; cassette VHS N° 19, cassette audio N° 19
 @Family of ANT: un grand frère, Jean+Baptiste (Bibo, Bébou, Bou, etc) ; une grande soeur, Julie (Uli)
 @Transcriber: Carine
 @Dependent : act, sit, com, pho
 @Comment : ce fichier a été fabriqué artificiellement à partir du début du fichier antoine19.cha

 @Time Start: 00:00:00
 *MAY: t' as fait un gros dodo ?
 *ANT: oui .
 *MAY: tu mets tes chaussons ?
 *ANT: oui .
 *MAY: d' accord .
 *ANT: je dis toujours voui [: oui], voui [: oui], voui [: oui], voui [: oui] .
 %act: Maya rit
 *MAY: tu vas jouer avec Clara aujourd'hui .
 *MAY: t' as envie ?
 *ANT: voui [: oui] , voui [: oui], je dis toujours voui [: oui], voui [:oui], voui [: oui] .
 %act: Maya rit
 *MAY: alors, tu trouves tes chaussons ?
 *MAY: c' est pas ça tes chaussons ?
 *MAY: regarde .
 *MAY: ou c' est à Jean+Baptiste ?
 *ANT: c' est trop petit .
 %com: MAY et ANT ont parlé en même temps
 *MAY: c' est à Jean+Baptiste, ça ?
 *ANT: non .
 *MAY: c' est à Bibo [: Jean+Baptiste] ou c' est à toi ?
 *ANT: nan [: non], c' est à Tato [: Antoine] .
 %pho: na~ sea tatø
 *MAY: c' est à Tato [: Antoine] ?
 *ANT: voui [: oui] .
 *MAY: et ben tu les mets , alors .
 *ANT: voui [: oui] .
 *ANT: là+haut .
 *MAY: on va en haut , d' accord .
 %sit: tous montent l' escalier
 *MAY: tu vas faire voir ta maison à Clara .
 *MAY: elle a pas vu encore .
 *MAY: tu vas lui faire voir ta chambre et tes affaires et la chambre de Julie, la chambre de papa et maman.
 *CLA: et moi, j' ai pas vu tu [: tout] .
 *CLA: j' ai que vu le salon, moi, et la cuisine .
 *CLA: j' ai que vu la cuisine maman .
 *MAY: alors .
 *CLA: le salon .
 *MAY: laisse, laisse le t' expliquer .
 *ANT: Clara è [: elle] dit maman [""] à Tato [: Antoine] .
 @Time Duration : 00:00:00-00:01:06
 @End

Transcrire à l'aide de l'éditeur de CLAN

L'éditeur (traitement de texte) de CLAN est spécialement conçu pour travailler avec les fichiers CHAT (.cha). Pour ouvrir la fenêtre de l'éditeur sous Windows, cliquer sur l'icône CLAN que vous aurez installée sur votre écran, ou double-cliquer sur un fichier avec l'extension .cha.

Vous pouvez utiliser l'éditeur pour d'autres tâches que la transcription, il fonctionne comme tout autre traitement de texte ASCII.

Pour transcrire ou travailler sur un fichier déjà transcrit, vous devez travailler en mode CHAT : cliquez sur *Mode*, puis sur *CHAT mode*. Si le fichier sur lequel vous travaillez possède l'extension .cha, vous serez automatiquement placé en *CHAT mode* lors de son ouverture. L'indication que vous êtes en *CHAT mode* est fournie par la barre en bas à gauche de votre écran : **CLAN [E] [CHAT] n** (numéro de la ligne où se trouve le curseur).

Les commandes fonctionnent de la même façon que dans la plupart des traitements de texte. Les menus de la barre de menu (*menu bar*), en haut de l'écran, qui sont les plus utilisés sont *File* (Fichier) et *Edit* (Edition). On trouvera sous le menu *File* les commandes habituelles : créer un nouveau fichier (*New*), ouvrir un fichier existant (*Open*), enregistrer (*Save*), enregistrer sous (*Save As*), aperçu avant impression (*Print Preview*), mise en page (*Print Preset*), suivies de la liste des derniers fichiers utilisés. Sous le menu *Edit*, les fonction usuelles sont disponibles : couper (*Cut*), copier (*Copy*), coller (*Paste*), rechercher (*Find*, *Find Same*), remplacer (*Replace*, *Replace and Find Next*). Une commande très utile du menu *Edit* est *Go To* qui permet d'aller sur une ligne précise en spécifiant son numéro.

Lorsqu'on travaille sur un fichier CHAT, l'éditeur prend en compte chaque participant listé dans le *Header @Participants*: et l'intègre dans le menu *Tiers* ("étages", lignes) de la barre de menu. Il suffit alors de cliquer sur *Tiers*, puis sur *Update* (mise à jour) et on peut alors entrer rapidement tous les débuts de ligne dans lesquels figurent les participants en cliquant sur leur codes. De même, on peut ajouter (au plus) quatre noms de lignes dépendantes en les listant sous le *Header @Dependent*: Exemple :

@Dependent: act, com, exp, sit

Ces lignes dépendantes pourront alors être entrées automatiquement en cliquant sur leur code dans le menu *Tiers*. Si on introduit des changements au niveau des participants ou des lignes dépendantes, il faudra à nouveau faire *Update* pour remettre à jour les codes disponibles.

Il est indispensable de vérifier fréquemment le fichier sur lequel on travaille (en lançant le programme CHECK sans sortir du fichier en cours de transcription) sinon on accumule les erreurs. Ceci est particulièrement important lorsqu'on débute dans la transcription en CHAT : il n'est pas possible d'apprendre à transcrire correctement si on ne corrige pas ses erreurs régulièrement. On lance CHECK en cliquant sur le menu *Mode*, puis sur *Check Opened File*. Une fois toutes les erreurs corrigées, sauvegarder le fichier en cliquant sur *Save* dans le menu *File* ou, plus rapidement, en cliquant sur la disquette en haut à gauche de l'écran (barre d'outils).

Fenêtre " Commandes " :

Lorsqu'on est dans l'éditeur du CHILDES, la fenêtre "Commandes" s'affiche en cliquant sur *Window* dans la barre de menu, puis sur *Commands*. Il faut alors choisir le répertoire sur lequel on désire travailler, c'est à dire celui dans lequel vos fichiers sont situés :

- 1) Cliquer sur le bouton *working* (répertoire de travail)
- 2) Localiser le répertoire qui contient les fichiers désirés
- 3) Cliquer sur le bouton *Select Current Directory* (sélectionner le répertoire en cours)

Vous êtes alors en mesure d'utiliser les commandes de la fenêtre *Commands*. Par défaut, le résultat de vos analyses ira dans le répertoire que vous avez choisi comme répertoire de travail ; si vous désirez qu'il aille dans un autre répertoire, vous devez spécifier ce dernier de la même façon en cliquant sur le bouton *output*, etc.

Tapez votre commande (dans le rectangle blanc), puis cliquez sur *Run* ou faites "Entrée" (pour un exemple de commande, cf. page 4). Pour voir les commandes qui sont à votre disposition, taper "info" dans la fenêtre commandes, puis cliquer sur *Run* ou faire "Entrée".

Le tableau de la page 12 énumère les principaux programmes CLAN avec une brève description.

Indications pour la transcription du français 19/07/2001 18:29:11

Quelques principes fondamentaux (rappel) :

Chaque ligne principale, commençant par un astérisque, (*) ne doit comporter qu'un seul énoncé, et doit être obligatoirement terminée par un signe de ponctuation. Les signes de ponctuation de base sont le point (.), le point d'interrogation (?) et le point d'exclamation (!) ; d'autres possibilités sont offertes par le CHILDES (Manuel CHAT, pp. 60-69). Pas de ponctuation à l'intérieur des énoncés, à l'exception de la virgule (,) et du point-virgule (;) ; jamais de points de suspension. Les pauses à l'intérieur d'un énoncé sont indiquées par des dièses (#) entourés d'espaces (# pour une pause simple, ## pour une pause plus longue, ### pour une très longue pause). Il est aussi possible de noter certains contours intonationnels sur la ligne principale, soit à l'intérieur d'un énoncé, soit en fin d'énoncé (Manuel CHAT, pp. 61-66). Lorsqu'un locuteur produit plusieurs énoncés à la suite, ils doivent faire l'objet de lignes principales distinctes. Donnons deux exemples de ponctuation en fin d'énoncé proposés par le système CHAT :

+... indique un énoncé incomplet, laissé en suspens par le locuteur

+/. est utilisé pour indiquer un énoncé interrompu par une autre personne ou un bruit

Les majuscules sont réservées aux lettres initiales des noms propres ; pas de majuscules en début de phrase, ni à l'intérieur des mots.

Les lignes dépendantes (commençant par le symbole %) doivent suivre immédiatement la ligne principale dont elles dépendent. On peut en mettre autant qu'on veut à la suite d'une ligne principale, pourvu qu'elles soient toutes différentes.

Lorsque ce que dit l'enfant est incompréhensible (à cause du bruit par exemple), on note le matériel non compris **xx** si on pense qu'il s'agit d'un seul mot, et **xxx** dans le cas où on considère qu'il peut y avoir plusieurs mots.

Lorsque les sons produits par l'enfant sont audibles, mais qu'on ne peut leur faire correspondre aucun mot, plusieurs solutions sont possibles. Si on a transcrit phonétiquement ces productions sur une ligne dépendante %pho utilisée systématiquement dans la transcription, on les notera, sur la ligne principale, **yy** (dans le cas où on pense qu'il s'agit d'un seul mot) et **yyy** (dans les autres cas). Sinon, on peut noter ces formes sur la ligne principale, en UNIBET (cf. tableau plus bas p. 13) et en les faisant précéder du signe **&** (ex : ouinouin sera noté **&wue~wue~**) ; ces productions pourront alors être soit ignorées, soit prises en considération selon les options des programmes CLAN. Une autre solution, qui peut être provisoire, consiste à noter ces productions avec l'orthographe du français, mais il faut être attentif au fait qu'elles seront traitées comme des mots par les programmes, et que l'orthographe française est loin d'être phonétique, ce qui entraîne que ni les notations, ni leurs interprétations ne sont univoques. En règle générale, il faut souligner que la ligne principale n'a pas vocation à transcrire la phonétique.

Lorsque certains locuteurs échangent des propos que l'on juge non pertinents pour l'étude, on ne transcrit pas ces paroles, mais on note **www** sur la ligne principale en la faisant suivre d'une ligne dépendante **%exp** dans laquelle on mentionne la raison de cette absence de transcription ; par exemple

***MOT:** **www .**

%exp: **MOT parle au téléphone**

Il est utile de connaître le temps que dure une session, ou de pouvoir retrouver un énoncé sur le magnétoscope à partir du compteur ; pour ceci, on fera systématiquement appel aux *headers* (cf. p. 2 du présent document) **@Time Start** (indiquant l'affichage du compteur au début de la séance) et **@Time Duration** (indiquant le temps qui sépare le début et celui du moment considéré), ex.

@Time Start: 00:00:00

@Time Duration: 00:00:00 01:12:00

De même, il est indispensable de pouvoir localiser les enregistrements correspondant à un fichier : cassettes vidéo, cassettes audio, etc. Tout d'abord, il convient d'apporter un soin tout particulier à l'étiquetage des cassettes ; de plus il ne faut jamais faire figurer plusieurs séances sur une même cassette. Tous les fichiers doivent comporter un *Header* (En-Tête) **@Tape Location**: spécifiant le détail des bandes correspondant à ce fichier. Exemple :

@Tape Location: **cassette vidéo 8mm N° 2 (original) ; cassette VHS Antoine 2 B (copie) ; cassette audio Antoine 2 C (copie)**

Les dates, les âges et les chiffres doivent être entrés selon les normes CHAT dans les *Headers* (En-Têtes). Par exemple, le 23 janvier 1995 devra être entré sous la forme **23-JAN-1995**. Les mois sont ainsi toujours en capitales, abrégés de la façon suivante : JAN, FEB, MAR, APR, MAY, JUN, JUL, AUG, SEP, OCT, NOV, DEC. En effet, les programmes utilisent ces informations pour calculer les âges. L'âge d'un sujet de 3 ans, 10 mois et 12 jours sera entré sous la forme **3;10.12**. On a vu plus haut les conventions adoptées pour entrer les chiffres correspondant au compteur du magnétoscope ou à l'heure du jour (**@Time Start**, **@Time Duration**).

Il est parfois utile de connaître la version CHAT utilisée pour la transcription lorsqu'on utilise des fichiers transcrits à diverses dates ; cette indication est donnée par le Header **@Coding** : . La version actuellement disponible sur le Web est celle du **11-JUL-2001** ; la version disponible sur le CD-ROM 2000 du CHILDES est celle du **11-NOV-1999**. Dans la mesure où CHAT est maintenant stable, cette information n'est utile que si on utilise des fichiers transcrits depuis longtemps et qu'on n'a pas repassé sous CHECK.

Instructions spécifiques au français :

Il faut être attentif au fait qu'en français, les deux points (:) sont précédés d'un espace alors que ce n'est jamais le cas dans les transcriptions. D'autres conventions typographiques propres au français ne s'appliquent pas non plus : pas de points de suspension, pas de guillemets, etc.

Les programmes CLAN reconnaissent les mots en tant qu'entités séparées par des espaces. Pour cette raison, nous ferons suivre les apostrophes d'un espace lorsqu'elles séparent deux mots différents (par exemple un article élide et un nom : *l' oiseau* ou un pronom élide et un verbe : *c' est* ; cf. exemples dans la transcription de la page 5 de ce document). On évitera les apostrophes à l'intérieur des mots ; ainsi on n'écrira pas *p'tit*, on utilisera d'autres solutions pour noter la prononciation (ligne dépendante phonétique %pho, parenthèses : *p(e)tit* ; remplacement : *ptit* [: *petit*], etc.) ; ces différentes solutions ne sont pas exclusives, et les options des programmes permettent de considérer toutes les possibilités. Cependant, certains mots français courants comportent une apostrophe interne (ex : *aujourd'hui*, *presqu'île*) et on peut aussi envisager de considérer comme un seul mot des expressions comme *d'accord* ou *d'abord* ; dans ces cas, et dans ces cas seulement, on ne mettra pas d'espace après l'apostrophe.

Le tiret qui sert en français à marquer, entre autres, le trait d'union dans les mots composés, est utilisé dans le CHILDES pour marquer les suffixes (exemples : *jardin-s*, *jardin-et*, *jardin-ier*, *jardin-er*, *je jardin-erai*). Ce codage des suffixes flexionnels et dérivationnels sur la ligne principale est optionnel. Il s'agit d'un codage morphologique superficiel qui n'est pas conseillé dans la mesure où il est difficile de faire un usage constant et systématique de ces marqueurs (CHAT Manuel, pp. 54-55) et que cela altère la lisibilité de la ligne principale ; de plus il nécessite qu'on ait préalablement listé les suffixes dans un fichier spécial. Etant donné que nous ne codons pas les affixes sur la ligne principale, il ne doit donc y avoir aucun tiret à l'intérieur des mots dans nos transcriptions. Nous utilisons, conformément au système CHAT, le signe plus (+) à la place du tiret français dans les mots composés : *grand+père*, *chewing+gum*. (Il est à noter que certains collègues étendent l'usage du symbole plus (+) à des locutions comme *parce+que*, ce qui présente l'avantage de les compter comme un seul mot ; de telles décisions peuvent faire l'objet de transformations systématiques sur les fichiers et ne doivent pas ralentir la transcription, qui suit sur ce point l'usage commun.)

Le tiret est aussi utilisé en français entre des mots différents pour marquer l'inversion verbe/pronom : *est-ce que, dit-il, vas-y, prends-en*, etc. Dans ces derniers cas, il sera remplacé par un espace dans les transcriptions, afin que les programmes puissent reconnaître les différents mots : *est ce que, dit il, vas y, prends en*, etc. Dans certains cas, les tirets entourent une lettre présente pour des raisons euphoniques et qui sert à éviter un hiatus (*voilà-t-il, a-t-il*, etc.) ; je propose de faire figurer cette lettre, qui est prononcée, entre parenthèses (*voilà (t) il, a (t) il*, etc.); ceci permet soit de la considérer, soit de l'ignorer selon les options choisies

Conventions adoptées dans nos transcriptions

Nous utilisons le système CHILDES tel qu'il est, en suivant les instructions de transcription telles qu'elles sont décrites dans le manuel CHAT (2000), et nous actualisons régulièrement les programmes CLAN, en les téléchargeant sur le site WEB du CHILDES dont l'adresse figure au début de ce document. Une contrainte minimale des fichiers ainsi transcrits est qu'ils doivent passer sans problèmes sous le programme de vérification CHECK. La seule addition que nous avons faite au système est l'adjonction d'un *Header* (En-Tête) **@Family of #:**. Cette convention permet de savoir quels sont les frères et sœurs de l'enfant étudié et sous quelles formes ils apparaissent dans les productions. Elle est spécifiée dans le fichier *00depadd.cut* (extension *.cut* pour CHILDES Utilities) qui doit figurer dans le même répertoire que les fichiers en cours de transcription ou d'analyse. Cette modification n'est pas essentielle et peut être supprimée sans dommage. Cependant elle permet de se familiariser avec l'utilisation du fichier *00depadd.cut*, qui pourra être nécessaire ultérieurement, lors du codage par exemple.

La transcription sur la ligne principale doit s'efforcer de respecter la façon dont les mots sont produits, les divergences avec les formes adultes, sans cependant aboutir à une multitude de variantes orthographiques préjudiciable aux analyses. Notons bien que nous parlons de "divergences" et non pas d' "erreurs". Une règle est de ne pas privilégier les solutions propres au français écrit ; c'est ainsi que la forme écrite *il y a* pourra, selon les cas, être transcrite *y a, i y a, ou il y a* selon la prononciation (nous ne mettons pas d'apostrophe après *y*).

Nous utilisons assez systématiquement la procédure de "*replacement*" (remplacement, Manuel CLAN, p. 72) pour distinguer la façon dont l'enfant produit un mot et sa forme adulte : le mot est donné sous la forme propre au locuteur considéré, immédiatement suivie de la forme standard entre crochets, précédées de deux points et d'un espace ; exemple **ptit [: petit]**. Les programmes CLAN peuvent alors analyser soit les formes "adultes" ou "normales" entre crochets, soit les formes effectivement produites, selon l'option choisie. Si la personne qui transcrit n'est pas sûre de la correspondance, on note différemment (exemple : **lalou [=? douche]**), ce qui indique qu'il s'agit d'une transcription douteuse. Les remplacements **[: texte]** ne doivent pas être confondus avec les autres symboles comportant des crochets []. Ils appartiennent bien à la même catégorie des *scoped symbols* (cf. plus bas), mais ils sont les seuls à être reconnus par les programmes en option ; les autres *scoped symbols*, comme **[=? texte]** sont des indications utiles à la lecture des transcriptions, mais ne peuvent pas être traités par les programmes à la place des mots effectivement prononcés par les sujets.

Les lignes dépendantes que nous utilisons le plus souvent sont :

- **%com:** qui permet d'introduire des commentaires à propos de l'énoncé qui précède ;
- **%act:** qui décrit les actions des participants concernés ;
- **%exp:** qui permet de préciser la nature (la position, etc.) des objets ou des personnes dont on parle ;
- **%sit:** qui permet de décrire certains aspects pertinents de la situation ;
- **%pho:** qui fournit une transcription phonétique en UNIBET ou en Alphabet Phonétique

Une catégorie importante de symboles du CHILDES est constituée par les *scoped symbols*, qui ont une portée correspondant à une partie d'un énoncé, voire à sa totalité (Manuel pp. 70-80) . Ces symboles sont inclus entre crochets [] [*square brackets*] et le matériel auquel ils réfèrent, lorsqu'il s'agit de plusieurs mots, est délimité par des parenthèses anguleuses < > [*angle brackets*] sur la ligne principale. Le matériel entre crochets fonctionne donc comme une description du matériel entre parenthèses pointues. Si un tel *scoped symbol* (symbole "à portée") ne porte que sur un seul mot, il n'est pas nécessaire d'entourer ce mot de parenthèses pointues. Nous avons vu un exemple de ces symboles plus haut avec le " remplacement ". Selon la nature du contenu des crochets, le matériel correspondant peut être inclus ou exclu de certaines analyses, comme le calcul du MLU. Les symboles de ce type que nous utilisons le plus fréquemment sont les suivants :

[: **texte**] remplacement : le texte entre crochets [] correspond à la forme standard adulte du mot ou de l'expression sur lequel il porte ; cette dernière expression doit être entourée de parenthèses pointues < > pour que les programmes puissent connaître ses limites précises ; exemples :

*ANT: <nan lu+là> [: pas celui+là] !

*ANT: ava [: au revoir] .

Dans ce dernier exemple, il n'est pas nécessaire d'utiliser des parenthèses pointues pour délimiter la portée (*scope*) du symbole de remplacement dans la mesure où celle-ci est restreinte à un seul mot.

[=? **texte**] le texte entre les crochets est une transcription alternative, ou une transcription douteuse ;

["] ce symbole marque une référence métalinguistique (citation) à un mot, à une partie de phrase ou à une phrase ; exemple :

*MAY: il a dit <nan lu+là> ["] .

[?] ce symbole indique que l'on n'est pas sûr de la transcription de la partie considérée ;

[/] reprise sans correction ; sauf indication contraire, la partie de la transcription considérée (répétée) est exclue de certaines analyses comme le MLU ;

[//] reprise avec correction ;

[+ **bch**] marque un énoncé à exclure de l'analyse (par exemple du calcul du MLU).

Tableau de divers symboles utiles avec leurs appellations anglaises et françaises

Symboles	Nom(s) Anglais	Nom(s) Français
	space	espace
-	dash, hyphen, soft hyphen, minus sign, subtraction	trait d'union, signe moins, tiret
=	equal sign, equal mark	signe égal
[]	square brackets	crochets
'	apostrophe, single quote, prime	apostrophe, prime
‘	inverted comma, single opening (or right) quote, forward quote	guillemet anglais simple, ouvrant
’	single closing (or left) quote, back quote	guillemet anglais simple, fermant
;	semicolon	point-virgule
/	(forward) slash, solidus	barre oblique, cotice, barre de fraction
\	backslash	barre oblique inversée, anti-slash, contre-cotice
.	period, dot, full stop, decimal point	Point
,	comma	virgule
!	exclamation mark	point d'exclamation
@	at sign, commercial at	arobe, <u>arrobas</u> , à commercial
#	number sign, ash, sharp, flatt	dièse, croisillon, symbole numéro
\$	dollar symbol	dollar
%	percent symbol	symbole "pour cent"
^	caret, turned V, insertion mark	lambda, chevron d'insertion
&	ampersand (<i>and per se and</i>)	esperluette, perluette, <i>et</i> commercial
*	asterisk, star	astérisque, étoile
()	round brackets, parentheses	parenthèses
<u> </u>	underline, underscore, underbar, horizontal bar	souligné, tiret de soulignement
+	plus sign, addition	signe plus
{ }	curly braces, curly brackets	accolades
"	quotation mark, double quote sign	guillemets anglais
~	tilde	tilde
:	colon	deux points
?	question mark, interrogation mark	point d'interrogation
	pipe, vertical bar	barre verticale
< >	angle brackets (less than sign, more than sign)	"parenthèses anguleuses", parenthèses pointues, (signe "inférieur à", signe "supérieur à")

Pour plus de détail sur les symboles utilisés en phonétique, voir tableau p. 13, et:

Pullum, Geoffrey K. & Ladusaw, William A. (1996). *Phonetic symbol guide*. Chicago : The University of Chicago Press

Principaux programmes CLAN et commandes

Il suffit de taper le nom d'un programme dans la fenêtre *Commands* (cf. présentation p. 6 et exemple p. 4) et de cliquer sur *Run* pour voir les options de ce programme s'afficher à l'écran. Les dernières commandes entrées sont disponibles en faisant défiler avec le curseur sur la barre de défilement (*scroll bar*) ; elles peuvent donc être effectuées à nouveau, avec éventuellement des modifications.

Programmes (Commandes)	Fonction
CHAINS	Analyse les codages des interactions entre locuteurs
CHECK	Programme de vérification des fichiers ; signale les erreurs
CHIP	Examine les répétitions et les expansions parent/enfant
CHSTRING	Effectue des changements de mots et de caractères dans les fichiers
COMBO	Cherche des combinaisons complexes dans les énoncés
COOCCUR	Cherche des patterns de co-occurrence entre mots
DATES	Calcule l'âge à partir de la date de naissance et de la date du jour
DIST	Examine les distances qui séparent les codages des actes de langage
DSS	Calcule le "Developmental Sentence Score" (Lee, 1974)
FREQ	Calcule les fréquences des mots dans un ou plusieurs fichiers
FREQMERG	Combine les résultats de plusieurs analyses faites avec FREQ
FREQPOS	Analyse les fréquences des mots selon leur positions dans l'énoncé
GEM	Permet de marquer certains passages pour des analyses spécifiques
GEMFREQ	Calcule les fréquences de mots à l'intérieur d'un passage marqué
GEMLIST	Liste les passages marqués par GEM dans un ou plusieurs fichiers
ID	Ajoute une ligne ID d'identification dans un fichier
KEYMAP	Liste les fréquences de codes qui suivent un code cible
KWAL	Cherche certains arrangements de mots et imprime la ligne
MAKEDATA	Convertit les fichiers (Windows, Macintosh, etc.) ; gère les fichiers
MAXWD	Trouve le mot le plus long d'un fichier
MLT	Calcule la longueur moyenne d'un tour de parole
MLU	Calcule la longueur moyenne des énoncés
MODREP	Compare la phonologie de l'enfant aux modèles parentaux
MOR	Insère une ligne %mor avec les catégories morphologiques de mots
PHONFREQ	Calcule les fréquences des phonèmes dans diverses positions
POST	"Part-of-speech tagger" ; Programme de désambiguïsation de la ligne %mor (cf. MOR) (Christophe Parisse, INSERM, Paris)
POSTLIST	S'utilise avec POST (")
POSTTRAIN	Elabore des fichiers utiles à POST pour la désambiguïsation (")
RELY	Mesure la fiabilité relative (<i>reliability</i>) entre deux transcriptions
SALTIN	Convertit des fichiers SALT en fichiers CHA
STATFREQ	Met en forme l'output de FREQ pour des analyses statistiques
TEXTIN	Convertit du texte normal au format CHAT
TIMEDUR	Calcule les recouvrements d'énoncés à partir du fichier audio
VOCD	Calcule l'indice de diversité lexicale VOCD (G. McKee, Reading)
WDLEN	Calcule les différentes longueurs d'énoncés en mots

Des commandes utilitaires, selon le vieux style DOS ou Unix, sont aussi disponibles

- **batch** [+ "nom de la macro" + *Run*) permet d'exécuter une macro existante ;
- **dir** (+ *Run*) affiche les fichiers du répertoire en cours,
- **ren** permet de renommer un ou plusieurs fichiers ; ex : **ren *.txt *.cha**
- **cd conv del** (dangereux !) **rmdir** (id) . **list tvne** etc

UNIBET et IPA

L'UNIBET permet de représenter les phonèmes à l'aide de caractères ASCII. Ceci n'est plus nécessaire, dans la mesure où l'éditeur du CHILDES peut maintenant utiliser les symboles de l'alphabet phonétique international (IPA). Le standard adopté est IPAPhon et on peut ajouter ces polices à partir du CD-ROM ou du site WEB du CHILDES dans le répertoire Fonts de Windows : c:\WINDOWS\FONTS. Cependant, l'utilisation des symboles IPA est plus longue et laborieuse, tout au moins au début et ces caractères ne doivent normalement être utilisés que dans la ligne dépendante %pho. De plus, certains corpus, comme celui de Grégoire, ont codé systématiquement la phonétique en symboles UNIBET. C'est la raison pour laquelle on trouvera ici le code UNIBET présenté dans le Manuel du CHILDES de 1995, p. 73, avec quelques erreurs corrigées. (dans le manuel 2000, une autre version, dite SAMPA UNIBET, diffère sur quelques points de celle-ci.)

Nous pensons que, à terme, les codages éventuels sur la ligne phonétique devraient se faire avec les symboles IPA. En attendant, nous proposons d'utiliser les symboles UNIBET tels qu'ils sont présentés dans le tableau ci-dessous, qui présente aussi les caractères IPA, lorsqu'il y aura lieu de coder la phonétique :

UNIBET	IPA	Orthographe	Exemples	UNIBET complet
a	a	a (e)	table, femme	tabl, fam
A	â	a (â)	pas, pâte	pA, pAt
6	\	e (ai)	cerise, faisan	s6riz, f6za~
e	e	é (er, ez, et)	parlé, parler, parlez, et	parle, e
E	é	è (ê, e, ei, ai, aî)	mètre, mettre, maître	mEtr
i	i	i (y)	rire, type	rir, tip
Y	y	u (û, eu)	mûr, dur, j'ai eu	mYr, dYr, ZeY
o	o	o (ô, ot, au, aud, eau)	mot, chaud	mo, So
O	ø	o (oo, au, u)	Rome, rhum, alcool, Paul	rOm, alkOl, pOl,
8	œ	eu (œu, uei)	fleur, bœuf, cueillir	f18r, b8f, k8jir
3	Ø	eu (œu)	feu, nœud	f3, n3
u	u	ou (oû, oux)	courir, goût	kurir, gu
a~	a~	an (am, en, em, ant, ent)	maman, vent	mama~, va~
e~	e~	in (ein, ain, int)	matin, main, plein	mAte~, me~, ple~
o~	o~	on, (ons, ont, om, omb)	selon, plomb	s6lo~, plo~
8~	œ~	un (um, eun)	un, parfum, à jeun	8~, parf8~, aZ1~

p	p	p (pp)	patin	pate~
b	b	b	balle	bal
m	m	m (mm)	maison	mezo~
t	t	t (tt, th)	tour	tur
d	d	d (dd)	dame	dam
n	n	n (nn)	nager	naZe
k	k	c (cc, k, qu)	carte	kart
g	"	g (gg, gu, c)	goût, second	gu, s6go~
f	f	f (f, ph)	carafe	karaf
v	v	v (w)	veux	v3
s	s	s (ss, sc, c, ç, ti)	si, ce	sî, s6
l	l	l (ll)	livre, aller	livr, ale
N	µ	gn	gagner	gaNe
r	Ë, ë	r (rr)	rue	rY
z	z	z (s)	gazon	gazo~
Z	Ω	g (ge, j)	plongeon	plo~Zo~
S	β	ch	chemin	Seme~
w	w	w (ou, o(i))	oiseau, jouet	wazo, Zue
W	¥	u (hu)	huile	Wil
j	j	i (ll, y)	miel	mjEl

Récapitulation de quelques symboles importants pour la transcription au format CHAT

Headers (En-Têtes) obligatoires

@Begin	doit obligatoirement figurer au commencement d'un fichier CHAT
@Participants:	liste obligatoire des participants (locuteurs) ; figure juste après @Begin
@End	doit obligatoirement figurer à la fin d'un fichier CHAT

Autres En-Têtes utilisés (CHI = Child)

@Age of CHI:	aa;mm.jj ; exemple 3;10.12 ; indique l'âge du locuteur CHI
@Birth of CHI:	jj-MMM-aaaa ; ex : 23-JAN-1995 ; indique la date de naissance du locuteur CHI
@Coding:	indique la version utilisée pour la transcription en CHAT ; exemple : 11-JUL-2001
@Comment:	commentaires
@Date:	jj-MMM-aaaa ; exemple : 05-DEC-1998 ; date de la séance
@Filename:	indique le nom du fichier
@New Episode:	indique l'endroit où un nouvel épisode commence et un autre s'achève
@Transcriber:	indique la personne qui a effectué la transcription
@Situation:	atmosphère générale et situation de la séance
@Tape Location:	indique sur quelles cassettes audio et vidéo se trouve la séance transcrite
@Time Start:	00:00:00 (exemple) temps initial (compteur du magnétoscope)
@Time Duration:	00:00:00-01:25:12 (exemple) temps écoulé depuis le moment initial
@Warning:	indique les défauts ou les limitations du fichier
@Family of CHI:	(rajouté), indique les membres de la famille de CHI et la forme sous laquelle ils apparaissent dans ses productions

Mots (sur la ligne principale)

xx	parole inintelligible, traitée comme un mot
xxx	parole inintelligible, traitée comme un groupe de mots
yy	inintelligible, transcrit dans la ligne %pho, traité comme un mot
yyy	inintelligible, transcrit dans %pho, traité comme un groupe de mots
www	matériel non transcrit
&	début de fragment phonologique sur la ligne principale (suivi de la transcription en UNIBET)
[?]	mot douteux, choix incertain ou meilleur choix
()	partie manquante d'un mot

Morphèmes

-	marqueur de suffixation (non utilisé)
+	marqueur des mots composés

Marqueurs prosodiques à l'intérieur des énoncés

#	pause entre les mots
-?	intonation montante

Ponctuation en fin d'énoncé

.	point
?	point d'exclamation
!	point d'interrogation
+...	énoncé non terminé, laissé en suspens
+/.	énoncé interrompu

Symboles "à portée" (scoped symbols) : <la portée> [le symbole]

["]	citation
[: texte]	texte : remplacement de la forme produite par le sujet par la forme standard
[= texte]	texte : explication sur la ligne principale
[=? texte]	texte : transcription alternative ou transcription douteuse
[/]	reprise sans correction
[//]	reprise avec correction
[+bch]	("back channel") énoncé exclu des analyses

Lignes dépendantes utilisées

%com:	permet d'introduire des commentaires à propos de l'énoncé qui précède ;
%act:	décrit les actions des participants concernés ;
%exp:	permet de préciser la nature (la position, etc.) des objets ou des personnes mentionnées dans l'énoncé précédent;
%sit:	permet de décrire certains aspects pertinents de la situation ;
%pho:	fournit une transcription phonétique (en UNIBET ou en Alphabet Phonétique International (IPA), cf. p. 13).

Les travaux que j'ai pu repérer sont les suivants (34 références) : (Bassano, 1996; Bassano, 1998a; Bassano, 1998b; Bassano, 2000; Bassano & Champaud, 1989; Bassano, Maillochon, & Eme, 1998; Bassano, Maillochon, Klampfer, & Dressler, 2001; Bassano & Mendes Maillochon, 1994; Bassano & Mendes Maillochon, 1995; Champaud, 1992; Champaud, 1993; Champaud, 1994; Champaud, 1995; Champaud, 1996a; Champaud, 1996b; Champaud, 2001; Champaud & Bassano, 1994; Champaud, Dressler, Gagarina, Sedlak, & Vollman, 1997; Ferdinand, 1994; Godard & Labelle, 1999; Hulk, 1996a; Hulk, 1996b; Klampfer, 2000; Kramer, 1993; Labelle, 1994; Labelle, 1996; Labelle & Valois, 1996; Parisse, 1999; Parisse & Le Normand, 1997; Parisse & Le Normand, 1998; Parisse & Le Normand, 2000a; Parisse & Le Normand, 2000b; Plunkett, 1999; Suppes, Smith, & Léveillé, 1973). Lorsqu'il s'agit de communications ou de manuscrits qui ont fait l'objet de publications ultérieures, seules ces dernières sont mentionnées.

- Bassano, D. (1996). Functional and formal constraints on the emergence of epistemic modality: A longitudinal study on French. *First Language*, 16(46, Pt 1), 77-113.
- Bassano, D. (1998a). L'élaboration du lexique précoce chez l'enfant français: structure et variabilité. (Structure and variability in French early lexical development). *Enfance*, 4, 123-153.
- Bassano, D. (1998b). Sémantique et syntaxe dans l'acquisition des classes de mots : l'exemple des noms et des verbes en français. *Langue Française*, 118(May), 26-48.
- Bassano, D. (2000). Early development of nouns and verbs in French: Exploring the interface between lexicon and grammar. *Journal of Child Language*, 27(3), 521-559.
- Bassano, D., & Champaud, C. (1989). The argumentative connective "même" in French: An experimental study in eight- to ten-year-old children. *Journal of Child Language*, 16(3), 643-664.
- Bassano, D., Maillochon, I., & Eme, E. (1998). Developmental changes and variability in the early lexicon: A study of French children's naturalistic productions. *Journal of Child Language*, 25(3), 493-531.
- Bassano, D., Maillochon, I., Klampfer, S., & Dressler, W. U. (2001). L'acquisition de la morphologie verbale à travers les langues (I). *Enfance*, 53(1), 81-99.
- Bassano, D., & Mendes Maillochon, I. (1994). Early grammatical and prosodic marking of utterance modality in French: A longitudinal case study. *Journal of Child Language*, 21(3), 649-675.
- Bassano, D., & Mendes Maillochon, I. (1995). L'émergence de la modalité de phrase dans le langage: de la prosodie à la grammaire. The acquisition of utterance modality: From prosody to grammar. *Enfance*, 2, 187-204.
- Champaud, C. (1992). *Early occurrences of tensed forms in French: The role of cognitive-semantic categories and utterance situation*. Paper presented at the Fifth Conference of the European Society for Cognitive Psychology, Paris, France, 12-16 September.
- Champaud, C. (1993). *Tense forms and functions in French-speaking children*. Paper presented at the Sixth International Congress for the Study of Child Language, Trieste, Italy, 18-24 July.
- Champaud, C. (1994). *The development of verb forms in French children around two years of age: some comparisons with Romance and non-Romance languages*. Paper presented at the First Lisbon Meeting on Child Language, Lisbon, Portugal, 14-17 June.
- Champaud, C. (1995). *Forms and functions of early verbs and verb modifiers in French-speaking children: A cross-linguistic perspective*. Unpublished manuscript, CNRS, Université Paris V, Paris.
- Champaud, C. (1996a). *Overgeneralization in early acquisition of verbal morphology: The case of French*. Paper presented at the Workshop: The Acquisition of Morphology in L1, 7th International Morphology Meeting, Vienna, Austria, 14-15 February.
- Champaud, C. (1996b). *Word order and dislocations in early French productions*. Paper presented at the VIIth International Congress for the Study of Child Language, Istanbul, Turkey, 14-19 July.
- Champaud, C. (2001, 28 mars 2001). *Une introduction au système CHILDES en français*, Toulouse, Université de Toulouse-Le Mirail.
- Champaud, C., & Bassano, D. (1994). French concessive connectives and argumentation: An experimental study in eight- to ten-year-old children. *Journal of Child Language*, 21(2), 415-438.
- Champaud, C., Dressler, W. U., Gagarina, N., Sedlak, M., & Vollman, R. (1997). *The early development of verb categories in French, German and Russian*. Paper presented at the VIIth European Conference on Developmental Psychology, Symposium: Comparative Study of Language Acquisition, Rennes, France, 3-7 September.

- Ferdinand, A. (1994). Semantic verb types and the acquisition of verb movement in French. *Proceedings - Eastern States Conference on Linguistics (ESCOL), 11*, 128-139.
- Godard, L., & Labelle, M. (1999). Utilisation des adverbes temporels par des enfants dysphasiques du primaire. *Revista Latina de Pensamiento y Lenguaje, 7*(2), 89-106.
- Hulk, A. (1996a). The delay inversion questions in the acquisition of French. In C. Parodi, C. Quicoli, M. Saltarelli, & M. L. Zubizarreta (Eds.), *Aspects of Romance linguistics: Selected papers from the Linguistic Symposium on Romance Languages XXIV March 10-13 1994* (pp. 241-252). Washington, DC: Georgetown University Press.
- Hulk, A. (1996b). The syntax of Wh-questions in child French. *Amsterdam Series in Child Language Development, 5*(68), 129-172.
- Klampffer, S. (2000, July 9-14). *On spatial concepts underlying abstract relations in language: evidence from non-spatial use of French, German and Russian spatial prepositions*. Paper presented at the Seventh International Pragmatics Conference, Special Topic Panel 'Spatial relations in natural language', Budapest.
- Kramer, I. (1993). The licensing of subjects in early child language. *MIT Working Papers in Linguistics, 19*, 197-212.
- Labelle, M. (1994). Acquisition de la valeur du temps passé par les enfants francophones. *Revue Québécoise de Linguistique, 23*(1), 99-121.
- Labelle, M. (1996). The acquisition of relative clauses: Movement or not movement? *Language Acquisition, 5*(2), 65-82.
- Labelle, M., & Valois, D. (1996). The status of post-verbal subjects in French child language. *Probus, 8*(1), 53-80.
- Parisse, C. (1999). Cognition and language acquisition in normal and autistic children. *Journal of Neurolinguistics, 12*(3-4), 247-269.
- Parisse, C., & Le Normand, M. T. (1997). Etude des catégories lexicales chez le jeune enfant à partir de deux ans à l'aide d'un traitement automatique de la morphosyntaxe. *Bulletin d'Audiophonologie, XIII*(6), 305-328.
- Parisse, C., & Le Normand, M. T. (1998). Traitement automatique de la morphosyntaxe chez le petit enfant. *Glossa, 61*, 22-29.
- Parisse, C., & Le Normand, M. T. (2000a). Automatic disambiguation of morphosyntax in spoken language corpora. *Behavior Research Methods, Instruments and Computers, 32*(3), 468-481.
- Parisse, C., & Le Normand, M. T. (2000b). How children build their morphosyntax: The case of French. *Journal of Child Language, 27*(2), 267-292.
- Plunkett, B. (1999). Targeting complex structures in French questions. *Proceedings of the Annual Boston University Conference on Language Development, 23*(2), 764-775.
- Suppes, P., Smith, R., & Léveillé, M. (1973). The French syntax of a child's noun phrases. *Archives de Psychologie, 42*(166), 207-269. (Un travail précurseur du CHILDES)