

Textométrie appliquée à des corpus d'interaction verbale entre un adulte et un enfant : exemples

Luigi Sansonetti

EA 2290 SYLED – Université de la Sorbonne Nouvelle – Paris 3 – 19 rue des Bernardins –
75005 Paris – France – luigi.sansonetti@wanadoo.fr

Abstract

Through the analysis of two adult-child dialogue corpora, lemmatized and categorized, we shall observe the relations between the uses of imperfect and simple past verb tense forms in narrative contexts. We shall apply textometric methods to observe the roles of the speakers and their mutual influence.

Résumé

À travers deux corpus, lemmatisés et catégorisés, d'interactions verbales entre un adulte et un enfant, nous observerons le rapport entre les emplois de l'imparfait et du passé simple en contexte narratif. Nous utiliserons la textométrie pour mettre en relief les rôles et influences des locuteurs.

Mots-clés : statistique textuelle, textométrie, analyse de corpus, acquisition du langage, interaction verbale adulte/enfant

1. Introduction

Dans le domaine de l'étude de l'acquisition du langage par l'enfant, nous construisons des corpus en retranscrivant des dialogues entre un adulte et un enfant. Dans notre exemple, un récit a été lu à l'enfant avant que la transcription du dialogue ne commence. Nous avons rassemblé plusieurs corpus¹ de ce type, faisant intervenir deux enfants différents (Julien et Mathilde) observés à des moments échelonnés dans le temps, en gardant pour toutes ces expériences le même livre illustré *Cricor* (Ungerer, 1958) comme support de l'entretien et le même adulte.

Les problèmes de l'acquisition du langage par l'enfant intéressent plusieurs communautés de chercheurs (psychologues, psycholinguistes, linguistes, cognitivistes...²). Dans un corpus longitudinal de plusieurs dialogues, nous pouvons observer des phénomènes de reprises et de reformulations de la part des deux locuteurs. Lorsque l'enfant reprend à l'adulte une construction et l'emploie dans une formulation inadéquate, et que l'adulte reformule cette construction de manière adéquate, l'adulte répond aux tâtonnements de l'enfant et lui permet alors de progresser dans la connaissance de la langue.

À l'aide de logiciels de statistique textuelle et d'étiquetage automatique, il est possible d'étudier des données orales et d'observer l'évolution des emplois du vocabulaire et de la

¹ Le premier corpus longitudinal, nommé « Julien », comprend trois dialogues entre une même adulte et un même enfant, entre 5 ans 10 mois et 6 ans 4 mois. Le deuxième corpus longitudinal, nommé « Mathilde », comprend aussi trois dialogues entre la même adulte et un autre enfant, entre 4 ans 9 mois et 4 ans 11 mois (Tissier, 2001).

² Nous pouvons citer les ouvrages de Martinot (2005), Marcos et al (2004), Hickmann (2003), Tomasello (2003), Lentin (1998), Blanche-Benveniste (1997).

syntaxe. Il est également possible, après des phases de préparation du corpus, de visualiser rapidement des reprises et des reformulations d'un locuteur à l'autre qui sont d'autant de phénomènes pertinents dans le processus d'acquisition de la langue en situation de dialogue.

Nous avons vu l'intérêt d'une approche lexicométrique dans l'analyse de corpus d'interactions verbales entre un adulte et un enfant, dans le cadre du repérage du phénomène des questions – réponses « pourquoi / parce que », des constructions syntaxiques de type « préposition suivie de verbe infinitif » (Sansonetti, 2004). Nous allons nous intéresser ici au temps de la narration dans le cadre d'une interaction triangulaire « livre – adulte – enfant ».

2. Préparation des corpus

Avant d'analyser les corpus, nous avons ajouté des « clés », permettant ainsi de partitionner le corpus longitudinal en dialogues ou en locuteurs (Lébart et Salem, 1994).

TRANSCRIPTION :	
A1	Alors c'est quoi l'histoire de <u>Crictor</u> ? J1 mm un jour le facteur arrive et donne un mm quand madame Bodot ouv(r)e le pa/quet mm elle va / elle va au zoo
A2	Et pourquoi elle va au zoo ? J2 pour voir si c'est pas un / un serpent dang(è)reux (il avale sa salive)

Tableau 1 : Extrait du corpus brut "julien"

Nous avons ensuite étiqueté ces corpus³ pour obtenir des versions catégorisées et lemmatisées. Nous gardons pour notre exemple les corpus avec l'étiquette adjointe à la forme lemmatisée. Les corpus « Julien » et « Mathilde » lemmatisés seront notés respectivement « ju-lemcat » et « ma-lemcat ».

<ad=101>	Alors c'est quoi l'histoire de <u>Crictor</u> ?
<ju=101>	mm un jour le facteur arrive et donne un mm quand madame Bodot ouv(r)e le pa/quet mm elle va / elle va au zoo
<ad=102>	Et pourquoi elle va au zoo ?
<ad=101>	alors_ADV ce_PDS être_VINDP3S quoi_PRI le_DETDFS histoire_NCFS de_PREP Crictor_NPI
<ju=101>	mm_INT un_DETIMS jour_NCMS le_DETDFS facteur_NCMS arriver_VINDP3S et_COO donner_VINDP3S un_DETIMS mm_INT quand_SUB madame_NCFIN Bodot_NPI ouvrir_VINDP3S le_DETDFS paquet_NCMS mm_INT elle_PPER3S aller_VINDP3S elle_PPER3S aller_VINDP3S au_DETDFS zoo_NCMS
<ad=102>	et_COO pourquoi_SUB elle_PPER3S aller_VINDP3S au_DETDFS zoo_NCMS

Tableau 2 : Extraits du corpus "julien" et de son codage (lemmatisation et catégorisation) par Cordial

3. Le corpus « ju-lemcat »

Lorsque nous regardons la répartition des énoncés par dialogue (Figure 1), nous remarquons que l'imparfait (en noir) est beaucoup plus utilisé que le passé simple (en gris) dans le 1er corpus (22 occurrences d'imparfait pour 8 occurrences de passé simple). Dans le 2ème corpus, il y a trois fois moins d'imparfait pour presque autant de passé simple (7 occurrences

³ Avec l'étiqueteur Cordial <http://www.synapse-fr.com/>

d'imparfait pour 6 occurrences de passé simple). La fréquence d'utilisation de chaque temps verbal suit ensuite une même progression dans le 3ème corpus (9 occurrences d'imparfait pour 7 occurrences de passé simple), mais l'imparfait reste globalement le plus utilisé.

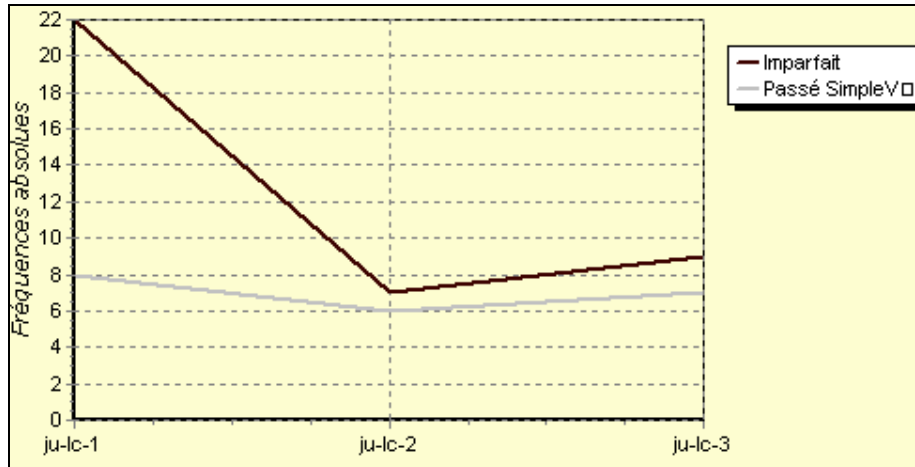


Figure 1 : Répartition des formes d'imparfait et de passé simple dans les dialogues (deux locuteurs confondus)⁴

L'examen des productions du locuteur (Figure 2) nous montre que c'est seulement dans le 1^{er} dialogue que l'adulte énonce plus d'imparfait que l'enfant (13 occurrences d'imparfait pour 9 occurrences de passé simple). Dans les deux autres dialogues, l'enfant en énonce plus que l'adulte (5 occurrences d'imparfait pour 2 occurrences de passé simple et 5 occurrences d'imparfait pour 4 occurrences de passé simple). Quant au passé simple, l'adulte en énonce 2 dans le 1^{er} dialogue et cesse par la suite. L'enfant, lui en énonce 6, 6 puis 7. Alors que l'adulte cesse d'utiliser le passé simple, l'enfant va en employer de plus en plus. L'adulte va favoriser l'imparfait (19 occurrences d'imparfait pour 2 occurrences de passé simple) alors que l'enfant va équilibrer l'emploi des deux temps (19 occurrences d'imparfait pour 19 occurrences de passé simple). À partir de cette observation, nous pouvons conclure que l'adulte reprend à l'imparfait les énoncés au passé simple de l'enfant.

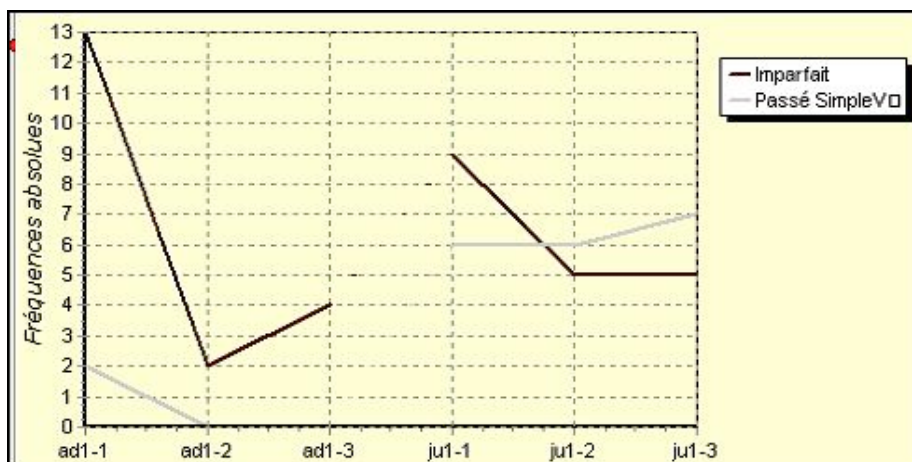


Figure 2 : Répartition des formes d'imparfait et de passé simple par locuteur⁵

⁴ « ju-lc-1 » correspond au 1^{er} dialogue, « ju-lc-2 » au 2^{ème} et « ju-lc-3 » au 3^{ème}.

À regarder de près (Figure 3), nous remarquons que dans les interactions, lorsque l'enfant emploie une forme de passé simple (en gris), il utilise parfois aussi de l'imparfait (en noir) (sur 10 énoncés contenant des formes d'imparfait, il y en a 4 contenant également des formes de passé simple notés en gris et noir en même temps).

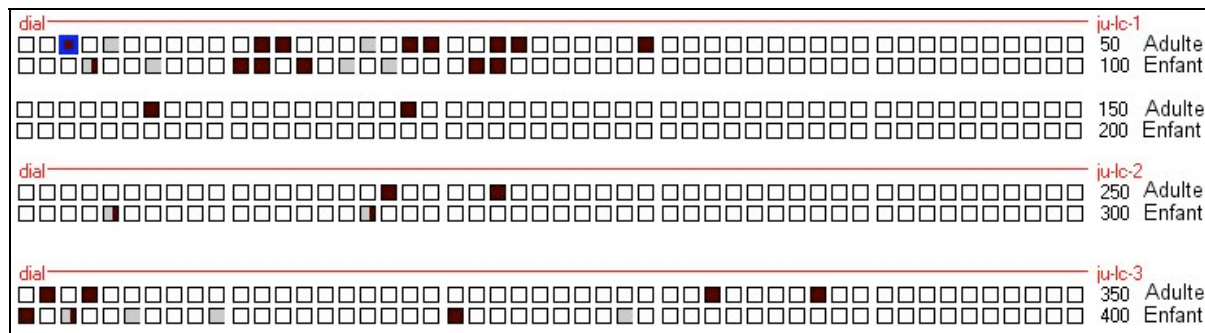


Figure 3 : Carte des énoncés contenant des formes d'imparfait et/ou de passé simple

Sur la figure 3, les interactions sont représentées sur deux lignes⁶. La ligne supérieure correspond aux énoncés de l'adulte, et la ligne inférieure à ceux de l'enfant. Chaque carré représente un énoncé et ce dernier se met en couleur s'il contient la forme recherchée.

Dans l'énoncé n°4⁷ du 1er dialogue, l'enfant va d'abord énoncer un imparfait (noté en gras), puis utiliser trois formes de passé simple (notées en gras italiques). L'adulte va reprendre immédiatement la forme de passé simple dans le même contexte lexical avec une modalité interrogative introduite par pourquoi. L'enfant fait ensuite deux tentatives d'emploi au passé simple (marquées CREA pour « création enfantine » dans la version étiquetée et en gras italique souligné dans la version brute) qui ne sont pas reprises ni reformulées par l'adulte, avec une production canonique.

J4 et mm un c'**était** un boa constructeur alors elle l'**appela** Cric(tor) elle lui **donna** un [s] / [s] le biberon [s] elle lui **apporta** des palmiers
 A5 Et pourquoi elle lui **apporta** des palmiers ?[...]
 J7 mm elle lui **faisa** un petit gilet,, elle **décida** de / de l'em/mener en classe il **apprena**., l'alphabet

Dans l'énoncé n°15, l'enfant produit un attacha au passé simple (en gras italique) qui sera immédiatement repris tel quel par l'adulte dans une phrase interrogative. L'enfant poursuit son récit avec une tentative de construction au présent (en gras italique souligné) qui se concrétise par « avoir » au passé simple. L'adulte reformule immédiatement ce verbe au présent⁸.

⁵ « ad1-1, ad1-2, ad1-3 » correspondent aux énoncés de l'adulte dans le 1^{er}, 2^{ème} et 3^{ème} dialogue, « ju1-1, ju1-2, ju1-3 » correspondent à ceux de Julien, pour avoir à gauche l'adulte et à droite l'enfant.

⁶ Pour obtenir cette représentation pratique à exploiter, nous avons soumis les transcriptions des corpus à une réorganisation : nous avons regroupé en tête du corpus les cinquante premiers énoncés de l'adulte suivis des cinquante énoncés produits par l'enfant en réponse.

⁷ Les conventions de transcription sont : entre parenthèses ce qui n'a pas été prononcé, entre crochet une indication phonétique, la barre oblique pour le tâtonnement, (ind.) pour un passage inaudible, les guillemets pour du texte lu et deux virgules collées pour marquer une pause courte.

⁸ Pour la lisibilité de nos exemples, nous présentons les extraits des corpus originaux. Selon les anciennes conventions de transcription, seuls les énoncés ne contenant que « mm » ne sont pas numérotés. C'est pourquoi,

J15 euh euh euh après le serpent l'*atta/cha* jusqu'à l'arrivée de la police
 A17 Il *attacha* qui le serpent ?
 [...]
 J17 alors il *a* / il / il pour sa récompense il *eut* une médaille une statue et son / un jardin
 A19 Donc i(l) *a* bien un jardin qui portait son nom ?

Dans le 2ème dialogue qui s'est déroulé 3 mois après, à l'énoncé n°5, l'enfant raconte l'histoire et utilise une forme de passé simple qui ne sera pas du tout reprise ni reformulée par l'adulte, avec d'autres verbes au présent et à l'imparfait. L'adulte continue par du présent.

Plus tard, à l'énoncé n°17, l'enfant finit de raconter l'histoire et utilise cinq formes au passé simple (en gras italique) avec une tentative (en gras italique souligné), tout en produisant également des présents (non marqués), des imparfaits (en gras) et un futur (non marqué). L'imparfait, immédiatement énoncé par l'adulte, n'est pas une reformulation d'un verbe mais n'est que la conclusion de l'histoire racontée par l'enfant.

J17 mm il / il *joua* avec les garçons il et aussi / il *joua* avec les filles il *était* très serviable il apprendra à faire les nœuds mm un jour le / le cambriolage se *passa* mm mm madame Bodot *était* déjà bâillonnée sur une chaise et ligotée lorsque le euh mm / le serpent se réveille mm il entou(re) il l'attache à la [f] / à l'arrivée de la police il *fut* / il / il *f(ut)* / on lui *mettait* une médaille et il se / il se / il *sculptait* et on / et on lui *metta* un jardin propre pour lui
 A18 Donc là c' *était* l'histoire de Crictor, le serpent.

Dans le 3ème dialogue à l'énoncé n°3, l'enfant produit quatre formes au passé simple et une forme du verbe être à l'imparfait. L'adulte reprend immédiatement le verbe être à l'imparfait dans un contexte lexical différent. À l'énoncé n°5, l'enfant continue le récit et utilise un passé simple canonique *mesura* après un autre non canonique *metta* qui n'est pas repris par l'adulte, ni reformuler.

À l'énoncé n°8, l'enfant produit un passé simple (en gras italique) que l'adulte reformule immédiatement au présent dans le même contexte lexical et syntaxique (en gras italique souligné).

J8 il va / il va / et là il va dans la neige et il / et madame Bodot *décida* de le prendre dans sa classe
 A11 Euh madame Bodot *décide* d(e) le prendre dans sa classe parce qu'elle est institutrice.

Le dernier passé simple à l'énoncé n°27 de l'enfant est une production spontanée qui n'est pas reprise par l'adulte par la suite.

Par ce retour au texte, nous remarquons que l'adulte ne reprend pas systématiquement le passé simple énoncé par l'enfant en mettant la forme à l'imparfait. D'autre part, l'analyse du texte du livre montre qu'il n'y a qu'un seul présent utilisé dans l'histoire. Les autres temps sont le passé simple et l'imparfait.

4. Le corpus « ma-lemcat »

Avec le corpus « ma-lemcat » nous allons pouvoir comparer, comme avec « ju-lemcat », l'emploi de l'imparfait par rapport au passé simple chez les deux locuteurs, mais aussi comparer les modes d'interaction de l'adulte avec le même support, avec des enfants différents et à 5 mois d'écart.

dans nos corpus informatisés, il peut y avoir un décalage dans la numérotation des énoncés. Les nouvelles conventions numérotent désormais tous les énoncés.

Lorsque nous comparons les fréquences d'emploi de l'imparfait par rapport au passé simple (Figure 4), nous constatons, à l'inverse de « ju-lemcat », que le passé simple est le plus utilisé sur l'ensemble des 3 dialogues (64 occurrences de passé simple en gris pour 56 occurrences d'imparfait en noir).

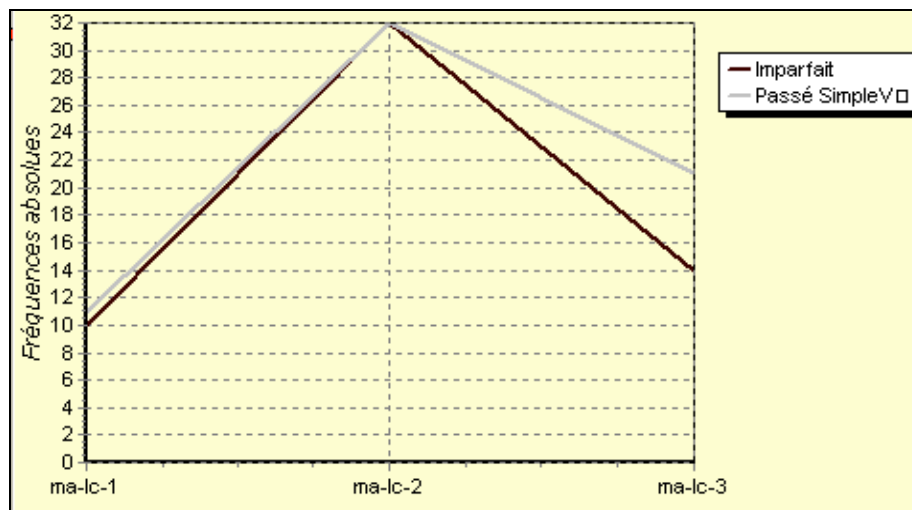


Figure 4 : Répartition des formes d'imparfait et de passé simple dans les dialogues (deux locuteurs confondus)

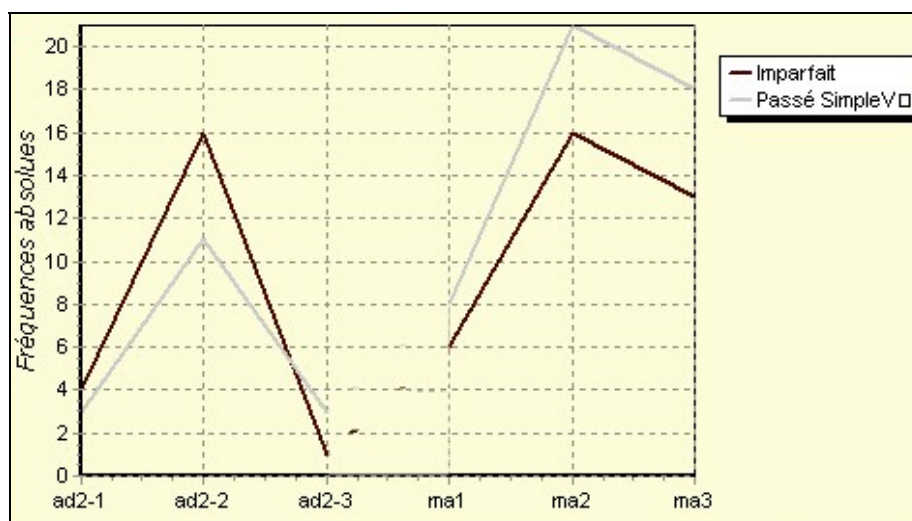


Figure 5 : Répartition des formes d'imparfait et de passé simple par locuteur

L'examen des productions du locuteur nous montre qu'il y a un rapport très étroit entre les évolutions d'emploi des deux temps pour l'enfant comme pour l'adulte (Figure 5) avec un pic d'occurrences dans le 2ème dialogue. L'enfant emploie toujours plus de formes au passé simple qu'à l'imparfait, alors que l'adulte ne le fait qu'au dernier dialogue. Enfin, alors que l'adulte emploie beaucoup moins de formes de passé simple et d'imparfait dans le dernier dialogue, l'enfant, quant à lui, a encore une forte utilisation de formes de passé simple et d'imparfait. Il nous faut vérifier, à partir de ce graphique, si l'adulte reprend les mêmes constructions énoncées par l'enfant, ou si c'est l'enfant qui reprend les formes employées par l'adulte.

Au regard de la carte des énoncés (Figure 6), nous avons confirmation d'un rapport plus étroit entre l'adulte et l'enfant dans l'emploi de l'imparfait (en noir) et du passé simple (en gris).

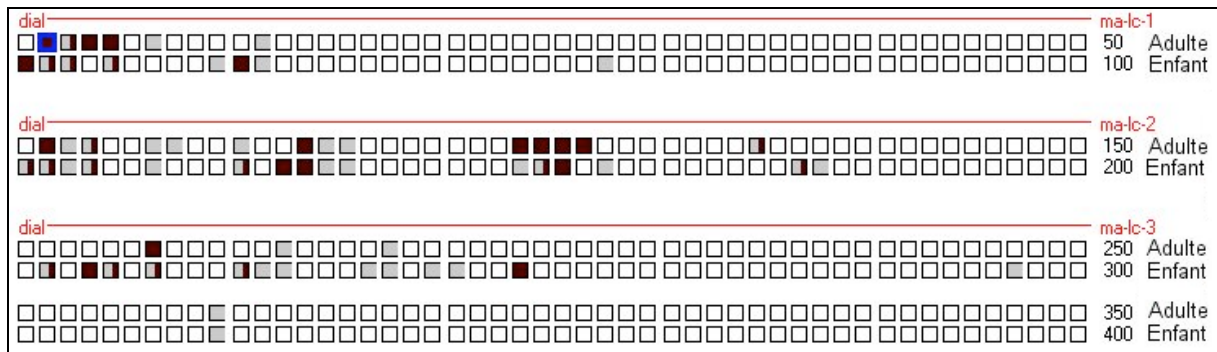


Figure 6 : Carte des énoncés contenant des formes d'imparfait et/ou de passé simple

À l'énoncé n°2 dans le 1er dialogue, l'enfant produit 2 formes au passé simple (en gras italique), une tentative de forme de passé simple du verbe ouvrir (en gras italique souligné) et un imparfait (en gras). L'adulte reprend immédiatement l'imparfait et reformule la tentative de l'enfant dans une forme canonique. L'enfant reprend immédiatement à son tour tout l'énoncé de l'adulte.

M2 Bodot et un monsieur lui *donna* des (ind.) il lui *acheta* un cadeau elle l'*ouvra* et elle *avait* (ind.) euh euh
 A3 Madame Bodot *avait* peur quand elle *ouvrit* la boîte.
 M3 quand / madame Bodot *avait* peur quand elle *ouvrit* la boîte

Peu après, à l'énoncé n°7, l'enfant produit 2 formes de donner au passé simple que l'adulte reprend dans le même contexte syntaxique et lexical avec changement de elle par madame Bodot. À l'énoncé n°10, l'enfant énonce un passé simple tricota immédiatement reformulé au présent par l'adulte. Puis à l'énoncé n°12, l'adulte produit une forme de dormir au passé simple que l'enfant reprend aussitôt.

Au début du 2ème dialogue qui s'est déroulé un mois après, il y a un 1er bloc de va-et-vient entre les temps sur les 8 premiers énoncés, avec des changements de temps parfois entre passé simple et présent, imparfait et passé composé. Le 2ème bloc est constitué de productions de formes de passé simple fut et donna par l'adulte, productions dues à la lecture du livre, qui sont reprises par l'enfant dans un autre contexte pour fut. Ce dernier ajoute dans le même énoncé un 3ème verbe amena au passé simple que l'adulte reprend à son tour.

A7 « Madame Bodot *fut* une vraie mère pour lui. Elle lui *donna* le biberon. »
 M7 oui,, madame Bodot *fut* fermière pour lui elle lui *donna* le biberon elle aussi / elle a lui / elle l'*amena* des palmiers pour euh remettre sa / pour
 A8 Madame Bodot *amena* des palmiers à Crictor pour que Crictor se rappelle de sa forêt natale.

Sur la fin du 2ème dialogue, il y a quelques emplois de passé simple qui ne sont pas repris ni reformulés.

Dans le dernier dialogue, nous remarquons comme une baisse des reprises de l'adulte des productions de passé simple de l'enfant. En effet, les productions de l'enfant sont, dans un premier temps, reformulées au présent. Ensuite, l'adulte ne reprend que le dernier verbe cité par l'enfant au passé simple, et ce à deux reprises dans le dialogue. À la fin, l'adulte énonce

spontanément un passé simple que l'enfant reprend immédiatement dans le même contexte puisqu'il s'agit d'une lecture d'un passage du livre par l'adulte.

5. Conclusion

Avec le corpus « ju-lemcat », nous avons émis l'hypothèse que l'adulte reformulerait par l'imparfait les formes de passé simple de l'enfant. Pour le corpus « ma-lemcat », nous pensions que les temps se répondaient de façon équilibrée puisqu'il y avait une même évolution des fréquences du passé simple et de l'imparfait. Par lectures successives des couches du corpus, nous avons pu infirmer certaines hypothèses et observer d'autres phénomènes.

Avec Julien, l'adulte n'a pas utilisé le passé simple, sauf dans la lecture du livre, survenue avant l'échange. Cela n'a pourtant pas empêché l'enfant d'en produire, mais sans réponse de la part de l'adulte, sans confirmation du bon usage ou d'une réutilisation possible dans d'autres contextes lexicaux et syntaxiques. L'adulte ne corrige pas, ne reprend que deux essais sur sept de formes de passé simple de l'enfant (*apprena* dans l'énoncé n°9 du 1er dialogue et *faisa* en énoncé n°28 du 3ème dialogue reformulés respectivement au passé composé *a appris* en énoncé n°11 du 1er dialogue et *a fait* en énoncé n°30 du 3ème dialogue). En revanche, l'adulte a bien suivi l'emploi des imparfaits par l'enfant. Avec la carte des énoncés (Figure 3), nous remarquons que c'est souvent suite à une énonciation de l'enfant que l'adulte produit un imparfait.

Avec Mathilde, nous avons un tout autre rapport entre l'adulte et l'enfant. L'adulte semble plus à l'écoute des propositions de l'enfant, tant à l'imparfait qu'au passé simple. Au point de produire, au 3ème dialogue, plus de passé simple que d'imparfait. D'autre part, l'adulte reformule une des deux tentatives de productions de formes de passé simple de l'enfant avec *ouvra*.

Pour les études à venir, nous nous intéresserons à deux observables. Quelques études sur le passé composé dans « ju-lemcat » montrent que l'emploi de ce temps est de plus en plus fréquent (1 puis 8 puis 13) par dialogue ainsi que par locuteurs. À l'inverse, dans « ma-lemcat », c'est un temps peu utilisé par rapport aux autres (6 occurrences de formes au passé composé puis 8 occurrences puis 9 occurrences). Nous pourrions, par l'exploration textuelle, observer la valeur aspectuelle ou temporelle de cette forme. La deuxième observable que nous traiterons prochainement, c'est la place du livre (et donc surtout du texte), par rapport aux énoncés des locuteurs. Nous aurons alors une analyse triangulaire plus fine entre le livre, l'adulte et l'enfant. Cette approche par textométrie permettrait d'observer le rôle des temps dans la narration et le rôle interactionnel de l'adulte dans le dialogue avec un enfant en cours d'apprentissage de la langue maternelle.

Références

- Blanche-Benveniste, C. (1997). *Approches de la langue parlée en français*. Paris, OPHRYS, Collection l'essentiel français.
- Hickmann, M. (2003). *Children's Discourse : Person, Space and Time across Languages*. Cambridge, Cambridge University Press
- Lébart L. et Salem A. (1994). *Statistique textuelle*. Dunod, Paris.
- Lentin, L. (1998). *Apprendre à penser, parler, lire, écrire*. Paris, ESF.
- Marcos, H. Salazar Orvig, A. Bernicot, J. Guidetti, M. Hudelot, C. Préneron, C. (2004) *Apprendre à parler : Influence du mode de garde*. Paris, L'Harmattan.

- Martinot, C. (2005). *Comment parlent les enfants de 6 ans ? Pour une linguistique de l'acquisition*. Besançon, Presses Universitaires Franc-Comtoises.
- Sansonetti L. (2004). Apports de la statistique textuelle pour le repérage des reprises et reformulations dans les corpus d'interactions verbales entre un adulte et un enfant. Actes du Colloque des 7èmes Journées internationales d'Analyse statistique de Données Textuelles, Louvain-la-Neuve, 10-12 Mars 2004, *Le poids des mots*, Purnelle, Fairon et Dister (éds.), 2 volumes, Presses Universitaires de Louvain : 993-999.
- Tissier C. (2001). *Rôle de l'adulte dans l'interaction langagière adulte-enfant (entre 4 ans 9 mois et 6 ans 4 mois) en situation de narrations dans deux corpus longitudinaux*. Paris, Mémoire de Maîtrise, ILPGA, Paris 3 (non publié).
- Tomasello, M. (2003). *Constructing a language : A usage-based theory of language acquisition*. Cambridge, Harvard University Press.
- Ungerer T. (1958). *CriCTOR*. L'École des Loisirs pour l'édition en langue française, 1980, réédition 2000. Collection « lutin poche ».

