HAL
open science

Anaphore et référence en production écrite : étude de textes narratifs d'élèves de 9 à 11 ans, du CE2 au CM2

Karine Peres Bonnemaison

▶ **To cite this version:**

Karine Peres Bonnemaison. Anaphore et référence en production écrite : étude de textes narratifs d'élèves de 9 à 11 ans, du CE2 au CM2. Linguistique. Université Toulouse le Mirail - Toulouse II, 2018. Français. NNT : 2018TOU20029 . tel-02627042

HAL Id: tel-02627042

https://theses.hal.science/tel-02627042v1

Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITE DE TOULOUSE

Délivré par :

Université Toulouse-Jean Jaurès

Présentée et soutenue par :
Karine PERES BONNEMAISON

Le jeudi 14 juin 2018

Titre :

ANAPHORE ET REFERENCE EN PRODUCTION ECRITE :
ETUDE DE TEXTES NARRATIFS D'ELEVES DE 9 A 11 ANS,
DU CE2 AU CM2

VOLUME 1

École doctorale et discipline ou spécialité :
ED CLESCO : Sciences du langage

Unité de recherche :
CLLE-ERSS UMR 5263

Directeur/trice(s) de Thèse :
Claudine GARCIA-DEBANC, Directrice
Josette REBEYROLLE, Co-Directrice

Jury :

M. Michel CHAROLLES, Professeur émérite, Université de Paris 3 Sorbonne Nouvelle

Mme Marie-Laure ELALOUF, Professeure, Université de Cergy-Pontoise, Rapporteur

Mme Claudine GARCIA-DEBANC, Professeure, Université de Toulouse Jean Jaurès, Directrice

Mme Josette REBEYROLLE, Maîtresse de Conférences, Université de Toulouse Jean Jaurès, Co-Directrice

Mme Marie-Noëlle ROUBAUD, Maîtresse de Conférences-HDR, Université d'Aix-Marseille, Rapporteur

REMERCIEMENTS

Je tiens à remercier ici tous ceux qui, de près ou de loin, de manière chronique ou plus ponctuelle, m'ont soutenue dans cette aventure un peu folle, il faut bien l'avouer, qu'est la présentation d'une thèse de doctorat.

Je remercie en premier mes directrices de recherche, Claudine Garcia-Debanc et Josette Rebeyrolle, qui m'ont fait confiance à l'issue de mon master d'enseignement et m'ont encadrée de manière exigeante tout au long de la thèse.

Merci à Michel Charolles d'avoir accepté d'être le président du jury, à mes rapporteurs, Marie-Laure Elalouf et Marie-Noëlle Roubaud pour leurs retours et de manière générale, à tous les membres du jury pour avoir accepté d'évaluer ce travail.

Merci aussi aux enseignants et enseignantes qui m'ont ouvert leur classe pour recueillir mes données, en particulier Jacqueline et Fanny, ainsi qu'à Mathilde Lala pour les textes transmis. Un grand merci également à Blandine pour la relecture finale et Claire et Jackie pour toute l'aide apportée.

Ma reconnaissance va également aux autres doctorants du laboratoire CLLE-ERSS (dont certains sont maintenant docteurs !), Marine L., Luce, Emilie M., Aleksandra, Karla, Marie-France, Maxime… Merci pour tous nos échanges, parfois entre deux couloirs, parfois autour d'un repas ou plus encore.

Je remercie particulièrement Florence, Karine B. et surtout Emilie C. (et tous ses post-its autour de mon écran d'ordinateur !) pour leurs encouragements et leur soutien lors des moments difficiles.

Cette thèse n'aurait pas pu voir le jour sans le soutien de tous mes amis, leur écoute et leur bienveillance malgré leur perplexité quant au monde de la recherche. Merci donc à Fanny, Roxane et Thomas, et en particulier à Jackie, Serge, Joël et Wendy.

J'adresse un remerciement tout particulier à un doctorant qui est aussi mon cousin, compagnon d'infortune dans les moments difficiles, malgré nos champs de recherche diamétralement opposés. Merci Manu pour nos échanges de mots et mails « remonte moral », édifiants sur tous les plans.

Bien sûr, je remercie ma famille pour son soutien moral et financier sans lequel je n'aurai pu mener cette thèse jusqu'à son terme. Merci aussi à ma belle-famille pour son intérêt manifesté tout au long de cette aventure.

Enfin, mes derniers remerciements vont à celui qui m'a soutenue dans les montagnes russes émotionnelles du quotidien et qui a réussi à tenir toutes ces années sans faillir. Merci Benoît parce que tu as toujours cru en moi et peut-être surtout quand moi je n'y arrivais plus.

Merci à cet Amour sans lequel rien n'aurait été possible.

« Ils ne savaient pas que c'était impossible, alors ils l'ont fait. »

Marc Twain.

INTRODUCTION **11**

PARTIE I : Les marques de cohésion à l'écrit : études linguistiques, psycholinguistiques et didactiques **19**

CHAPITRE 1 : Recherches linguistiques sur les notions de cohérence et de cohésion .. **21**

1.1. La cohérence : une notion difficile à définir **21**
1.1.1. L'approche cognitive : la cohérence en compréhension *vs* en production 22
1.1.2. L'approche linguistique : les quatre méta-règles de cohérence textuelle, selon Charolles (1978) 26
1.2. Le statut informationnel **30**
1.2.1. L'information nouvelle *vs* l'information ancienne 30
1.2.2. L'impact du statut informationnel sur le choix des formes linguistiques : l'accessibilité des référents 32
1.2.3. La progression thématique 35
1.2.4. La notion de saillance 36
1.3. La cohésion textuelle selon Halliday & Hasan (1976) **39**
1.3.1. La substitution 40
1.3.2. L'ellipse 41
1.3.3. La conjonction 42
1.3.4. La cohésion lexicale 43
1.4. La continuité référentielle **45**
1.4.1. La description et le traitement des expressions référentielles 45
1.4.1.1. La référence personnelle 47
1.4.1.2. La référence démonstrative 48
1.4.2. Quelles formes linguistiques pour aider à construire la référence ? 50
1.4.2.1. Les noms propres 50
1.4.2.2. Les déterminants des syntagmes nominaux 51
1.4.2.3. Les pronoms personnels 55
Synthèse des notions centrales du chapitre 1 ..... **57**

CHAPITRE 2 : Anaphores et Chaînes de Référence **59**

2.1. L'anaphore **59**
2.1.1. L'anaphore : une approche plurielle 60
2.1.1.1. L'anaphore comme phénomène textuel 60
2.1.1.2. L'anaphore comme phénomène mémoriel 66
2.1.1.3. L'anaphore : un phénomène complexe 69
2.1.2. Les différentes catégories d'anaphore 70
2.1.2.1 Les anaphores nominales 70
2.1.2.2. Les anaphores pronominales 73
2.2. Les chaînes de référence **74**
2.2.1. Les chaîne de référence : définition 75
2.2.2. Les relations entre chaînes de référence 79
2.2.3. Les chaînes de référence et les genres textuels 81
2.2.4. Les chaînes de référence et le genre narratif 83
2.2.5. Les chaînes de référence et la construction des personnages dans les textes narratifs 86

SYNTHESE des notions centrales du chapitre 2 ..... **87**

CHAPITRE 3 : Aspect développemental : acquisition des marques anaphoriques dans les productions orales et écrites d'enfants en fonction de leur âge **89**

3.1. Les marques anaphoriques en compréhension et en production **89**
3.1.1. Le traitement des marques anaphoriques en compréhension 91
3.1.2. Les marques anaphoriques en production 94
3.1.2.1. L'introduction des référents dans la narration : une maîtrise tardive du SN indéfini en position initiale 96
3.1.2.2. Le maintien des référents dans la narration : une construction progressive avec le pronom pour expression privilégiée 97
3.1.2.3. La gestion de la cohésion textuelle en production écrite 100
3.2. Les facteurs influençant les stratégies anaphoriques **107**
3.2.1. Effet du type de texte sur les procédés anaphoriques utilisés 107
3.2.2. Effets de la connaissance partagée sur l'utilisation des expressions référentielles 108
3.3. La place de la gestion de la cohésion textuelle dans les processus rédactionnels **110**
Synthèse des notions centrales du chapitre 3 ..... **113**

CHAPITRE 4 : L'enseignement des marques de cohésion et le jugement de cohérence des enseignants **115**

4.1. L'enseignement des marques de cohésion **115**
4.1.1. La cohérence textuelle dans les programmes d'enseignement 116
4.1.1.1. De la grammaire traditionnelle à la grammaire rénovée du *Plan de Rénovation* 116
4.1.1.2. L'émergence de la grammaire textuelle dans les programmes français de l'école primaire 120
4.1.1.3. Le retour de la grammaire phrastique traditionnelle et compromis dans les programmes français publiés en 2008 126
4.1.1.4. Les programmes d'enseignement de 2015 : une prise en compte équilibrée des « trois grammaires » ? 129
4.1.1.5. La place de la cohérence/cohésion dans les programmes d'enseignement de la grammaire de quelques pays francophones 133
4.1.2. L'enseignement des marqueurs référentiels 137
4.2. Évaluer la cohérence et la cohésion textuelles : la place de la cohésion et de la cohérence dans l'évaluation et le jugement de cohérence des enseignants **140**
4.2.1. L'évaluation des productions écrites d'élèves 141
4.2.2. Les enseignants et le jugement de cohérence 150
SYNTHESE des notions centrales du chapitre 4 ..... **152**

PARTIE II : Mise en œuvre de l'étude et méthodologie **155**

CHAPITRE 5 : De l'analyse de textes d'élèves à la constitution de grands corpus de textes scolaires **157**

5.1. Les corpus de productions écrites d'apprenants en français langue première : vers une meilleure représentativité **158**
5.1.1. Les enjeux de la constitution des corpus d'écrits scolaires 158

5.1.2. Les quelques tentatives de constitution de grands corpus d'écrits scolaires en français 163
5.2. Quelques difficultés pour la constitution et le traitement de corpus d'écrits scolaires 166
5.2.1. Des difficultés épistémologiques et méthodologiques 167
5.2.2. La transcription : une étape incontournable ? 169
5.3. L'exploitation de corpus.....173
5.3.1. Quelques outils d'aide à l'annotation 173
5.3.2. Les corpus constitués en vue d'étudier la référence: un manque de corpus d'écrits scolaires 175
SYNTHESE des notions centrales du chapitre 5.....177

CHAPITRE 6 : Le recueil de données et la constitution du corpus d'étude.....181

6.1. Problématique et questions de recherche 181
6.2. Recueil des données.....184
6.2.1. Protocole expérimental 184
6.2.2. Public choisi et conditions de recueil..... 186
6.3. Méthodologie de travail pour la transcription des textes du corpus 189
6.3.1. Transcription diplomatique..... 189
6.3.1.1. Codage des opérations d'écriture : ajout, suppression, remplacement, déplacement et ratures 190
6.3.1.2. Codage de la mise en page..... 193
6.3.1.3. Récapitulatif des conventions de codage utilisées pour la transcription diplomatique 196
6.3.2. Transcription normée 196
6.3.2.1. Normalisation orthographique 197
6.3.2.2. Normalisation ponctuationnelle 203

CHAPITRE 7 : Critères pour l'analyse de la référence dans les textes d'élèves et problèmes d'interprétation.....213

7.1. Catégories d'annotation 214
7.1.1. Informations générales à l'échelle du texte entier 215
7.1.2. Forme des mentions du référent..... 216
7.1.3. Modifieurs des mentions du référent..... 218
7.1.4. Fonctions syntaxiques des mentions du référent 219
7.2. Problèmes d'interprétation référentielle 220
7.2.1. Ambiguïté référentielle : plusieurs candidats à la référence 221
7.2.2. Non perception des contraintes morphosyntaxiques..... 223
7.2.3. Contradiction entre les marques syntaxiques du genre et la référence attendue 225
7.2.4. Problème d'accessibilité du référent dans le choix des formes linguistiques des mentions 227
7.2.5. Inférence avec calcul interprétatif coûteux..... 229
7.2.6. Discordance sémantique 230

PARTIE III : Résultats de l'étude 233

Chapitre 8 : Introduction et maintien du référent du pronom personnel anaphorique *elle* 235

8.1. Introduction du référent du pronom personnel *elle*.....235
8.1.1. Étude des formes de la première mention du référent du pronom personnel *elle*..... 237
8.1.2. Étude des expansions dans la première mention du référent du pronom personnel *elle* 243
8.1.3. Étude des fonctions syntaxiques des formes de M1..... 251
8.2. Maintien du référent du pronom personnel *elle* 253
8.2.1. Vue d'ensemble des CDR incluant le pronom personnel *elle* 253
8.2.2. Différentes configurations des CDR 255
8.2.2.1. Étude des cas d'anaphore ou de cataphore : 1 mention + P1 255
8.2.2.2. Étude des CDR minimales : 2 mentions + P1 257
8.2.2.3. Étude des CDR plus longues : 3 mentions (ou plus) + P1..... 260
8.3. Bilan de l'étude du référent du pronom personnel *elle*..... 263

CHAPITRE 9 : Introduction et maintien du référent du pronom personnel *il*.....265

9.1. Introduction du référent du pronom personnel *il*.....265
9.1.1. Etude des formes de la première mention du référent du pronom personnel *il* 268
9.1.2. Etude des expansions de la première mention du référent du pronom personnel *il* . 275
9.1.3. Étude des fonctions syntaxiques des formes de M1..... 280
9.2. Maintien du référent du pronom personnel *il*.....282
9.2.1. Vue d'ensemble des CDR du pronom personnel *il* 282
9.2.2. Différentes configurations des CDR 284
9.2.2.1. Étude des cas d'anaphore ou de cataphore : 1 mention + P2 284
9.2.2.2. Étude des CDR minimales : 2 mentions + P2 286
9.2.2.3. Étude des CDR plus longues : 3 mentions (ou plus) + P2..... 289
9.3. Bilan de l'étude du référent du pronom personnel *il* 293

CHAPITRE 10 : Introduction et maintien du référent du SN démonstratif *cette maison*

.....295

10.1. Introduction du référent du SN démonstratif *cette maison* 296
10.1.1. Etude des formes de la première mention du référent du SN démonstratif *cette maison* 297
10.1.2. Etude des expansions de la première mention du référent du SN démonstratif *cette maison* 303
10.1.3. Etude des fonctions syntaxiques de la première mention du référent du SN démonstratif *cette maison*..... 308
10.2. Maintien du référent du SN démonstratif *cette maison*.....311
10.2.1. Vue d'ensemble des CDR incluant le SN démonstratif *cette maison* 311
10.2.2. Différentes configurations des CDR 313
10.3. Bilan de l'étude du référent du SN démonstratif *cette maison* 320

CHAPITRE 11 : Introduction et maintien du référent du SN démonstratif *ce grand bruit*

.....323

11.1. Introduction du référent du SN démonstratif *ce grand bruit* 324
11.1.1. Etude des formes de la première mention du référent du SN démonstratif *ce grand bruit*..... 326

11.1.2. Etude des expansions de la première mention du référent du SN démonstratif *ce grand bruit* 331
11.1.3. Etude des fonctions syntaxiques de la première mention du référent du SN démonstratif *ce grand bruit.* 334
11.2. Maintien du référent du SN démonstratif *ce grand bruit* 335
11.2.1. Vue d'ensemble des CDR incluant le SN démonstratif *ce grand bruit* 336
11.2.2. Différentes configurations des CDR 337
11.3. Bilan de l'étude du référent du SN démonstratif *ce grand bruit*.....346

CHAPITRE 12 : Enchâssement des référents de la troisième phrase imposée : *Depuis cette aventure, les enfants ne sortent plus la nuit* dans l'ensemble du texte 349

12.1. Le SN défini *les enfants* 350
12.1.1. Introduction du référent du SN défini *les enfants* 350
12.1.1.1. Étude des formes de la première mention du référent du SN défini *les enfants*..... 356
12.1.1.2. Etude des expansions de la première mention du référent du SN défini *les enfants* 361
12.1.1.3. Etude des fonctions syntaxiques de la première mention du référent du SN défini *les enfants* 362
12.1.2. Maintien du référent du SN défini *les enfants* 364
12.1.2.1. Vue d'ensemble des CDR incluant le SN défini *les enfants* 364
12.1.2.2. Différentes configurations des CDR incluant le SN défini *les enfants* 365
12.2. Le SN défini *la nuit*.....372
12.2.1. Introduction du référent du SN défini *la nuit*..... 372
12.2.2. Maintien du référent du SN défini *la nuit* 381
12.3. Enchâssement de P3 et résolution de l'anaphore résomptive *depuis cette aventure* 383

CHAPITRE 13 : Bilan des résultats et comparaison 389

13.1. Quels sont les procédés d'introduction d'un nouveau référent dans un récit utilisés dans les textes étudiés ?.....390
13.1.1. Introduction des personnages : les référents des pronoms personnels clitiques de troisième personne *elle* et *il*, le SN défini *les enfants* 390
13.1.2. Introduction des entités non humaines, non animées : les SN démonstratifs *cette maison* et *ce grand bruit*, l'indication temporelle *la nuit* 393
13.2. Quels sont les procédés de coréférence employés dans les textes étudiés ? 395
13.2.1. Anaphores ou CDR ? Nombre moyen de mentions..... 396
13.2.2. Compositions des CDR des entités de la consigne 399
13.3. Les expressions linguistiques de la consigne font-elles toutes l'objet d'au moins une mention dans le texte, mise à part celle de la phrase imposée ?.....400
13.3.1. Les constituants de la phrase *fermoir* P3..... 402
13.3.2. Les référents des SN démonstratifs *cette maison* et *ce grand bruit*..... 404
13.3.3. Les référents humains des pronoms personnels clitiques de troisième personne *elle* et *il*..... 404
13.3.4. Nombre d'occurrences introduites dans les textes 405

CONCLUSION.....409

Liste des Figures 421

Liste des Tableaux **425**

INDEX **429**

BIBLIOGRAPHIE **431**

INTRODUCTION

Écrire un texte, c'est construire un ensemble de phrases dont l'objectif est d'être interprété et compris par un lecteur. En d'autres termes, un texte intelligible est un texte dans lequel s'articulent des idées en un rapport de sens logique. Un texte possède donc des qualités et attributs qui lui confèrent une certaine *cohérence*. De manière très générale, la cohérence peut être appréhendée comme la fabrique du sens d'un texte (Beaudet, 2001). Dans cette perspective, la cohérence prend en compte le sens global du texte par la considération à la fois d'opérations inférentielles et de marques linguistiques formelles (Charolles, 1995).

Dans la mesure où « dans un texte chaque phrase est interprétée par rapport à certaines autres phrases » (Van Dijk, 1977, cité par Beaudet, 2001), ces marques linguistiques formelles permettent de relier des énoncés. Depuis les travaux fondateurs d'Halliday & Hasan (1976), ces marques de relation entre énoncés ou constituants d'énoncés sont regroupées sous la notion de *cohésion textuelle*. La cohésion textuelle dépend de diverses marques de cohésion, tels que temps verbaux, connecteurs, marques de la référence. Comme l'explique Bronckart (1996 : 123) : « fondamentalement articulées à la linéarité du texte, [les marques de cohésion] en explicitent, à l'intention du destinataire, les grandes articulations hiérarchiques, logiques et/ou temporelles ».

Dans cette thèse, nous avons choisi de nous intéresser à un des aspects de la cohésion textuelle : la référence. Afin qu'une suite de phrases puisse être perçue comme un texte, il faut qu'elle présente une continuité référentielle minimale. Les expressions responsables de cette continuité sont connues sous le nom d'*anaphores* (Apothéloz, 1995a).

L'origine de ce choix remonte à nos travaux du Master Enseignement. Par la fréquentation régulière des élèves, nous avions remarqué leurs difficultés à construire des textes cohérents. Dans notre mémoire, qui traitait de la gestion des personnages dans des productions écrites narratives de cycle 3, nous nous sommes intéressées à la question de la continuité référentielle en appuyant notre analyse sur un corpus de textes d'élèves déjà collectés. Dans notre thèse, il nous semblait donc naturel d'approfondir cette étude, d'une part en affinant l'analyse et d'autre part en élargissant le corpus.

En tant que marques cohésives, la référence et la continuité référentielle, en particulier les phénomènes d'anaphore, ont fait l'objet de nombreuses recherches du point de vue de l'analyse et de la réception chez le locuteur adulte (on citera par exemple les travaux suivants : Charolles, 1978 ; Reichler-Béguelin, 1988a ; Kleiber, 1994 ; Apothéloz, 1995a ; Corblin, 1995 ; Cornish, 2009), dont le principal objectif est de décrire et comprendre leur fonctionnement. En revanche, l'étude de la référence du point de vue de la production, et particulièrement chez les enfants, ne concerne qu'un petit nombre de travaux (Charolles, 1988c ; Schnedecker, 1995 ; Masseron & Schnedecker, 1995). Bien que décrivant les modalités de l'expression référentielle chez les enfants, ces travaux s'attachent à l'étude « instantanée » (Schnedecker, 1995) d'un petit échantillon de textes.

Or, si l'on veut déterminer des indices de connaissance et de compétences des élèves tout au long du cursus scolaire, il est nécessaire d'adopter une démarche de progressivité qui étudie les compétences en fonction du niveau de classe, à l'image des travaux sur l'acquisition de l'orthographe et du contrôle orthographique tout au long des trois dernières années de l'école primaire (Geoffre, 2013). Toutefois, peu nombreux sont les travaux en didactique du français langue première qui se placent de ce point de vue (Nonnon, 2010). Une des raisons qui peut expliquer la rareté des travaux dans une perspective de progressivité est sans doute la faible quantité de données disponibles. En effet, l'analyse de l'acquisition de compétences sur plusieurs niveaux de classe nécessite des données en nombre suffisant pour conduire des études précises qui mettent en évidence des invariants et des éléments de progressivité (Garcia-Debanc & Bonnemaison, 2014). L'étude doit donc s'appuyer sur de grands corpus de textes scolaires. Or, d'une part, la constitution et le traitement de tels type de corpus est source de difficultés propres aux textes d'élèves, et d'autre part, les « grands » corpus de textes scolaires sont très peu nombreux car dans la tradition des travaux en didactique du français langue première, l'analyse de textes scolaires s'attache généralement à une analyse fine d'un échantillon très limité de textes (Fabre-Cols, 2000 ; Masseron, 2005 ; Boré, 2011).

L'acquisition des marques de la référence a fait l'objet de nombreuses études dans le domaine de la psycholinguistique, bien que celles-ci portent plutôt sur la réception que sur la production. Les travaux menés sur l'acquisition des marques anaphoriques du point de vue de la production ont pour objectif d'apporter des indications relatives à l'âge auquel les différentes unités et structures de la langue apparaissent dans les productions orales ou écrites des enfants. Les travaux de Karmiloff-Smith notamment (1979, 1980, 1981, 1983), ont permis de mettre au jour différentes étapes dans le développement de la référence chez les enfants

entre 3 et 10 ans. Pour le français, les recherches conduites par Maya Hickmann (1980, 1984, 1987, 1995, 2000) ont montré les procédés utilisés par les enfants entre 6 et 11 ans pour introduire et maintenir des référents dans le discours. Néanmoins, la plupart des travaux s'appuient sur l'étude de référents au caractère humain/non humain ou animé/non animé, laissant de côté les référents à d'autres constituants des énoncés. En outre, dans les recherches portant sur le développement des phénomènes anaphoriques en production, les travaux concernant la production écrite sont moins nombreux que ceux portant sur la production orale (Bronckart & Schneuwly, 1991 ; De Weck, 1991 ; De Weck & Schneuwly, 1994).

Face à ces constats, l'enjeu de notre étude est de mener des investigations qui permettent de décrire, comprendre et expliquer l'acquisition des marques de la référence en production écrite chez les élèves de différents niveaux de classe, concernant à la fois référents humains, évènement et indications spatio-temporelles. L'objectif de notre travail de thèse est, en d'autres termes, de proposer une cartographie des marques de la référence privilégiées par les élèves, en fonction d'une part, de leur niveau scolaire, mais également des caractéristiques des référents. Ces éléments permettront de mettre au jour des indicateurs de compétence des élèves concernant la référence et la cohésion textuelle. Sur le plan didactique, ces indicateurs devraient conduire à une meilleure compréhension de la maîtrise de la référence et de la cohérence textuelle par les élèves et pourront d'une part, servir de support à l'évaluation et au jugement de cohérence des enseignants, et d'autre part, permettre d'identifier des points d'entrée pour l'intervention didactique.

Un autre des enjeux de notre thèse est la contribution à la constitution d'un « grand » corpus de textes d'élèves en français langue maternelle, de différents niveaux de classe, conçu pour analyser la référence chez les élèves et dont le traitement par une double transcription (diplomatique et normée) lui permettra d'être mis à la disposition de la communauté scientifique.

Dans cette perspective, nous reprenons une tâche-problème conçue par C. Garcia-Debanc que nous avons proposée à 210 élèves du CE2 au CM2. La consigne d'écriture est inspirée de Charolles (1988c), mais s'appuie sur un ensemble de phrases plutôt que sur une bande dessinée. La tâche-problème proposée consiste à raconter une histoire dans laquelle les élèves doivent insérer trois phrases imposées. Ces trois phrases comportent des marques linguistiques nécessitant l'introduction de référents aux caractéristiques diverses.

Notre étude se propose de répondre aux questions suivantes :

1) Au niveau de l'introduction des référents : tous les référents font-ils l'objet d'une mention dans le texte, en dehors de celle contenue dans la phrase imposée ? Quelles sont les caractéristiques des expressions référentielles employées en première mention ?

2) Au niveau du maintien de la référence : pour référer, les élèves utilisent-ils des anaphores (2 mentions) ou bien des chaînes de référence (3 mentions et +) ? Quelle est la composition des chaînes de référence (longueur moyenne, type d'expressions référentielles) ?

Ainsi, notre étude cherche à déterminer dans quelles mesures les caractéristiques des référents (humain/évènement/indication spatio-temporelle) influent sur le choix des formes linguistiques utilisées pour les introduire et les maintenir dans le discours en fonction du niveau de classe, et dans quelle mesure les problèmes de continuité référentielle affectent la cohérence textuelle.

Ces questions ont pour arrière-plan théorique un ensemble de recherches relevant de plusieurs paradigmes. D'une part, les travaux en linguistique concernant la cohérence et la cohésion textuelles. Au niveau intra et inter-phrastique, Halliday & Hasan (1976) ont décrit les différentes marques cohésives qui permettent de lier les énoncés entre eux dans les textes : ellipse, substitution, conjonction, cohésion lexicale et référence. Pour tenter de décrire des règles de bonne formation des textes, Charolles (1978) a déterminé 4 méta-règles de cohérence : répétition, progression, non contradiction et relation. Notre travail s'intéressant particulièrement à la référence, il prend plus précisément appui sur les travaux qui décrivent les phénomènes d'anaphore (Corblin, 1985 ; Kleiber, 1988 ; Reichler-Béguelin, 1988a ; Apothéloz, 1995a) et les chaînes de référence (Corblin, 1987a ; Schnedecker, 1997).

D'autre part, notre étude repose également sur les travaux en psycholinguistique concernant le développement et l'acquisition des marques anaphoriques chez les enfants (Karmiloff-Smith, 1979 ; Hickman, 1987 ; De Weck, 1991 ; Jisa, 2000 ; Berman, 2004).

Enfin, notre travail s'intéressant aussi aux notions de cohérence et de cohésion textuelles d'un point de vue didactique, nous appuyons notre étude sur les travaux en didactique portant sur l'enseignement des marques de cohésion (Tauveron, 1987 ; Charolles, 1988a ; Reichler-Béguelin, 1995) et le jugement de cohérence des enseignants (Halté, 1989 ; Rondelli, 2008 ; Masseron, 2011).

La première partie de cette étude, constituée de quatre chapitres, est consacrée à la présentation du cadre théorique dans lequel nous nous inscrivons. Dans le chapitre 1, intitulé « Recherches linguistiques sur les notions de cohérence et de cohésion », nous tentons de définir les notions de cohérence et de cohésion textuelles, en lien avec le statut informationnel des référents. Dans le chapitre 2, « Anaphores et Chaîne de Référence », nous décrivons la notion d'anaphore à travers différentes approches (textuelle ou mémorielle) ainsi que celle de chaîne de référence (CDR) en lien avec divers genres textuels. Dans le chapitre 3, « Aspect développemental : acquisition des marques anaphoriques dans les productions orales et écrites d'enfants en fonction de leur âge », nous présentons une revue des travaux qui étudient l'acquisition des marques de la référence chez les enfants ainsi que leur gestion de la cohésion textuelle en production écrite. Enfin, pour clôturer cet état de l'art, dans le chapitre 4 « L'enseignement des marques de cohésion et le jugement de cohérence des enseignants », nous nous intéressons à l'aspect didactique de la cohérence et de la cohésion textuelles. Dans ce dernier chapitre, nous examinons premièrement les modalités de mise en œuvre de l'enseignement de ces notions par le croisement entre les directives présentes dans les programmes d'enseignement officiels et les enseignements proposés dans les classes. Deuxièmement, nous nous interrogeons sur la place de la cohésion dans l'évaluation ainsi que sur les enjeux du jugement de cohérence des enseignants. Dans chacun de ces chapitres, après un panorama non exhaustif des travaux éclairant nos différents axes de recherche, nous précisons ce que nous retenons dans le cadre de notre étude.

C'est au regard de ces éléments théoriques que nous présentons, dans la seconde partie de notre étude, sa mise en œuvre ainsi que les outils méthodologiques mobilisés pour le recueil et l'analyse de nos données. Dans le chapitre 5 « De l'analyse de textes d'élèves à la constitution de grands corpus de textes scolaires », nous présentons les enjeux de la constitution des corpus d'écrits scolaires et les difficultés qu'ils soulèvent, en faisant une revue de questions des points méthodologiques formulés dans les études précédentes sur la constitution de tels corpus. Nous montrons en particulier qu'il n'existe pas, à l'heure actuelle, de corpus de textes d'élèves regroupant plusieurs niveaux de classe constitué dans le but d'analyser la cohérence textuelle et plus particulièrement la référence. Dans le chapitre 6 « Le recueil de données et la constitution du corpus d'étude », nous présentons les questions de recherche qui ont présidé à la constitution de notre corpus d'étude et indiquons les modalités du recueil de nos données. Dans la deuxième partie de ce chapitre 6, nous présentons notre méthodologie de travail pour la transcription des textes du corpus. Enfin, dans le chapitre 7

« Critères pour l'analyse de la référence dans les textes d'élèves et problèmes d'interprétation », nous indiquons les critères retenus pour l'analyse de nos données et discutons de cas limites pour lesquels se posent des problèmes d'interprétation.

La troisième partie de notre étude expose et analyse les résultats obtenus. Dans la tâche-problème proposée, les phrases de la consigne contiennent des occurrences qui doivent faire l'objet d'une introduction. Au total, six occurrences sont analysées dans notre étude. Chaque chapitre présente l'analyse d'une occurrence de la consigne. Cependant, les occurrences ne sont pas analysées en suivant leur ordre dans les phrases de la consigne car nous avons fait le choix d'analyser à la suite les occurrences présentant les mêmes caractéristiques : les référents des pronoms personnels *elle* et *il*, les référents des SN démonstratifs *cette maison* et *ce grand bruit*.

Ainsi, le chapitre 8 intitulé « Introduction et maintien du référent du pronom personnel anaphorique *elle* » présente, comme son nom l'indique, les résultats de l'analyse concernant l'introduction et le maintien du référent du pronom personnel *elle*, qui est la première occurrence de la consigne. Le chapitre 9 « Introduction et maintien du référent du pronom personnel *il* » présente les résultats obtenus suite à l'analyse du référent du personnage masculin. Les deux chapitres suivants, c'est-à-dire le chapitre 10 « Introduction et maintien du référent du SN démonstratif *cette maison* » et le chapitre 11 « Introduction et maintien du référent du SN démonstratif *ce grand bruit* », traitent les résultats obtenus après l'analyse des occurrences de la consigne contenant un SN démonstratif. Le chapitre 12 quant à lui, intitulé « Enchâssement des référents de la troisième phrase imposée : *Depuis cette aventure, les enfants ne sortent plus la nuit* dans l'ensemble du texte », regroupe les occurrences de la troisième phrase de la consigne car cette dernière phrase a un statut particulier par rapport aux autres. Ce chapitre présente donc à la fois les résultats obtenus suite à l'analyse de chacune des occurrences de la dernière phrase, mais également la manière dont la phrase entière est intégrée au texte.

Tous les chapitres de la partie III « Résultats de l'étude » sont constitués selon la même architecture afin de pouvoir être lus séparément, mais également de permettre au lecteur de s'appuyer sur une structure récurrente qui donne un cadre régulier à la présentation des résultats. Ainsi, les chapitres 8 à 12 exposent, pour chaque occurrence à traiter, premièrement les procédés mis en œuvre pour introduire un nouveau référent et ensuite, les procédés mis en œuvre pour le maintenir dans le discours.

La partie III se clôt avec le chapitre 13, intitulé « Bilan des résultats et comparaison ». Ce dernier chapitre propose un état des lieux des principaux résultats de notre recherche en prenant appui sur les questions de recherche développées dans la partie II. D'autre part, ce chapitre 13 présente une analyse plus globale du traitement des occurrences de la consigne en les mettant en regard les unes des autres afin de déterminer les différences de traitement.

À la fin de notre exposé, un index regroupe les termes correspondants aux concepts les plus centraux pour l'étude. La référence indiquée renvoie à la page dans laquelle se trouve la définition du terme. D'autre part, la thèse est également constituée d'un volume d'annexes contenant les versions scannées des textes d'élèves de notre corpus.

PARTIE I :

Les marques de cohésion à l'écrit : études linguistiques, psycholinguistiques et didactiques

CHAPITRE 1 : Recherches linguistiques sur les notions de cohérence et de cohésion

La notion de cohérence textuelle peut être décrite, dans une première approximation, comme l'explicitation de ce qui constitue une norme minimale de composition textuelle. Que ce soit à l'échelle de la phrase ou bien à celle du texte, on peut penser qu'il existe un ensemble de règles qui régissent la bonne formation d'un discours, car comme le dit Charolles (1978 : 8) : « Comme tout tas de mots ne donne pas une phrase, tout tas de phrases ne forme pas un texte ».

Un discours (oral ou écrit) peut être considéré comme une succession d'énoncés qui entretiennent tous entre eux des rapports de sens (Charolles, 2011). Ces rapports de sens rendent l'enchaînement des énoncés cohérent et par là-même, donnent au texte entier sa cohérence. Cependant, l'étude de la cohérence ne se limite pas à la description des rapports de sens interagissant lors de l'interprétation d'un discours. Pour mieux cerner son fonctionnement, il faut aussi s'intéresser à la notion de cohésion textuelle ainsi qu'au statut de l'information véhiculée à travers les énoncés.

C'est ce que nous nous proposons de faire dans ce chapitre. En 1.1., nous tentons de décrire la cohérence des points de vue cognitif et linguistique, en tenant compte de l'articulation entre compréhension (du côté du lecteur) et production (du côté du scripteur). En 1.2., nous exposons un autre des principes relatifs à la cohérence : le statut informationnel des référents. Nous l'appréhendons à travers le prisme de la saillance linguistique, facteur de cohérence et de cohésion d'un énoncé, et plus particulièrement, à travers la distinction entre information nouvelle et information ancienne. En 1.3., nous étudions la notion de cohésion afin de montrer en quoi elle est intimement liée à celle de cohérence. Pour clore ce chapitre, en 1.4., nous nous intéressons aux marqueurs de la continuité référentielle en décrivant de manière non exhaustive des formes linguistiques qui marquent la référence.

1.1. La cohérence : une notion difficile à définir

La cohérence textuelle a fait l'objet d'abondants travaux et tentatives définitoires. Étudiée sous divers angles, notamment linguistique, psycholinguistique ou cognitif et définie

différemment selon les auteurs, la cohérence est de fait une notion difficile à cerner, d'autant plus qu'elle est associée à la notion de cohésion.

La distinction entre la cohérence et la cohésion d'un texte n'est pas toujours très nette parce que non seulement ces deux notions entretiennent des liens étroits (on s'appuie sur l'une pour décrire l'autre et vice-versa), mais elles sont aussi subordonnées l'une à l'autre. La question qui se pose est toutefois de savoir dans quel ordre. D'ailleurs, comme le fait remarquer Jaubert (2005), le terme de *cohesion* en anglais est quasiment le seul utilisé, en témoignent les travaux de Halliday & Hasan (1976), alors que *coherence* ou *coherency* relèvent plutôt de la philosophie. En français, le choix de l'un ou l'autre de ces termes correspond à des prises de position théoriques : Lundquist (1980) par exemple, privilégie le concept de *cohérence* alors que Bronckart & *al.* (1985) met en avant les phénomènes de *cohésion.* Toutefois, une définition large fait à peu près consensus, comme l'explique Charolles (1988b : 53) :

> « […] tout le monde est à peu près d'accord pour opposer d'un côté la cohérence, qui a à voir avec l'interprétabilité des textes, et, de l'autre les marques de relation entre énoncés ou constituants d'énoncés. Concernant ces marques, depuis M.A.K Halliday et R. Hasan (1976) on tend à les regrouper sous le nom générique de cohésion ».

Ainsi, la distinction est faite entre les deux termes. D'une part, la cohérence est une construction du lecteur qui relève de l'interprétation, de la réception, de l'accessibilité de l'information (Berrendonner & *al*., 2016). D'autre part, la cohésion concerne la construction interne du discours et relève de l'ensemble des marques linguistiques qui assurent les liens intra- et interphrastiques.

A la lumière de cette distinction, nous laissons de côté pour le moment la notion de cohésion (nous y reviendrons en détail dans la section 1.3) pour tenter de décrire de plus près celle de cohérence, en l'examinant d'abord du point de vue cognitif (1.1.1.) puis du point de vue linguistique (1.1.2.).

1.1.1. L'approche cognitive : la cohérence en compréhension *vs* en production

La cohérence textuelle peut s'étudier aujourd'hui sous divers angles. Cela est lié au fait que le texte est l'objet d'étude de nombreux domaines de recherche, comme par exemple la

pragmatique du discours, l'analyse textuelle des discours, la sémantique textuelle, la psycholinguistique textuelle, l'interactionnisme socio-discursif. Nous ne tentons pas ici une synthèse exhaustive des différentes approches, mais nous reprenons succinctement quelques éléments du bilan dressé par Rondelli (2010a) afin de mieux comprendre comment la cohérence est définie en fonction du domaine envisagé. Après ce bilan, nous présentons la définition que nous privilégions dans notre propre recherche.

L'analyse textuelle des discours considère que chaque texte est agencé au moyen de régularités qui lui donnent la propriété de former une unité cohésive et cohérente. Ces régularités dépendent des actions langagières interactives associées aux activités sociales et donnent naissance à différents genres de textes (Bronckart, 1997). Les textes sont vus comme des « produits verbaux résultant de la tension entre une situation d'énonciation singulière et les régularités intertextuelles d'un genre » (Rondelli, 2010a : 57). Partant, chaque texte est constitué par une architecture interne d'unités clairement identifiables qui s'articulent entre elles par différents types de liage, comme le montre Adam (2005). Ce sont ces liages qui donnent au texte sa cohérence.

Dans les théories interprétatives, le texte est considéré comme un ensemble constitué par des normes (Rastier, 2001). Les unités textuelles existantes sont décrites d'un point de vue morphosyntaxique mais ne sont pas considérées comme des unités constitutives *en soi* du sens d'un texte. La construction de la cohérence devient alors un parcours interprétatif, guidé par la connaissance du contexte et prenant appui sur des configurations qui forment une unité de sens du fait de l'ensemble du texte.

Pour les cognitivistes et psycholinguistes, la cohérence d'un texte se construit par des opérations entre le connu, le déjà lu et le nouveau « en train d'être lu ». Cette approche nous intéresse particulièrement. En effet, la cohérence n'est pas une notion seulement linguistique, comme l'explique Charolles (1995 : 9) :

> « […] la reconnaissance de ce qui rend un discours cohérent implique non seulement l'interprétation des éventuelles marques de cohésion qu'il comporte mais encore, et beaucoup plus fondamentalement, la mise en œuvre d'opérations inférentielles et, singulièrement d'inférences de liaison, portant conjointement sur le contenu du donné discursif, la situation dans laquelle il est communiqué et les connaissances d'arrière-plan des sujets. »

La cohérence concerne donc également « l'interprétabilité des textes » (Charolles, 1988b). Elle est en lien avec les représentations cognitives associées à la structuration des

textes car elle se fonde sur les aspects pragmatiques de leur fonctionnement. La cohérence « […] is a matter of semantic and pragmatic relations in the text »[1] (Reinhart, 1980 : 163). Ainsi définie, la cohérence « résulte de la capacité du destinataire à fonder son texte sur un savoir partagé » (Apothéloz, 1995a : 11).

Selon les travaux en psycholinguistique de Mazur-Palandre (2009), s'appuyant notamment sur les recherches de Favart (2005), lors de la production d'un texte, l'encodeur réfère à des situations, des personnages, des lieux, etc. qu'il doit encoder de manière à permettre au destinataire de se créer une représentation mentale cohérente du message. Cette représentation mentale est appelée, selon les auteurs, *modèle mental* (Johnson-Laird, 1980, 1983), *modèle de situation* (Van Dijk & Kintsch, 1983) ou *scénario* (Sanford & Garrod, 1981a, 1981b). L'intégration d'un message écrit en un contenu de sens s'organise sur plusieurs niveaux de représentation mentale : le code de surface, la base de texte, le modèle de situation, le type de texte et le niveau de communication (Graesser & *al.*, 2003, Chuy & Rondelli, 2010).

Il nous semble important de nous attarder un instant sur la description de ce modèle, en nous appuyant sur Chuy & Rondelli (2010). La notion de représentation mentale permet en effet, selon nous, de mieux comprendre les enjeux de la compréhension d'un texte et nous amène à comprendre le fonctionnement de sa production.

Les trois premiers niveaux de représentation mentale - *code de surface, base de texte, modèle de situation* - ont été identifiés par Van Dijk & Kintsch (1983) et sont les fondements de la compréhension d'un texte.

Le *code de surface* comporte les informations syntaxiques et lexicales telles qu'elles apparaissent dans le texte et sans lesquelles il ne pourrait y avoir construction de la représentation mentale. C'est le premier niveau que le lecteur doit décoder lorsqu'il est confronté à un texte.

La *base de texte*, le deuxième niveau, correspond au découpage de l'information en unités sémantiques correspondant à des propositions du texte. Leur construction par le lecteur repose donc uniquement sur l'information sémantique, sans que la formulation ou la syntaxe ne soient prises en compte. De cette manière, la structure des micro-propositions est régie par la cohérence locale car celle-ci détermine comment une unité textuelle contribue à la

[1] « […] est une question de relations sémantiques et pragmatiques dans le texte. » (notre traduction)

compréhension de l'unité précédente. La structure des macro-propositions est régie par la cohérence globale car elle montre comment l'unité textuelle contribue au développement du thème général.

Le troisième niveau est celui du *modèle de situation*. Le modèle de situation est créé par l'interaction entre les propositions explicites (*base de texte*) et les connaissances stockées en mémoire à long terme. La mémoire à long terme est un « système cognitif de capacité illimitée, permettant de stocker les informations pour plusieurs jours, voire plusieurs décennies » (Chuy & Rondelli, 2010 : 85 ; voir aussi Gaonac'h & Larigauderie, 2000). Le modèle de situation correspond à l'univers du texte. Ce niveau permet au lecteur de combler des lacunes laissées par le texte grâce aux processus d'inférence, d'intégration des connaissances antérieures et des savoirs partagés, afin de créer une représentation mentale plus complète et plus cohérente. La représentation mentale ainsi créée est mise en mémoire et le référent peut être récupéré le moment venu sous la forme d'un anaphorique par exemple. C'est ce que Reichler-Beguelin (1988a ; 1989) appelle la *mémoire discursive,* (cf. chapitre 2 section 2.1.1.2.).

Après les trois niveaux précédemment cités, deux autres niveaux jouent un rôle dans la construction de la cohérence : celui du *type de texte* et celui de la *communication* (Sperber & Wilson, 1989 ; Graesser & *al*. 2003).

Sans présenter en détail les différentes typologies des chercheurs, nous voulons indiquer ici que le niveau du *type de texte* se rapporte à la classe à laquelle appartient le texte (narratif, argumentatif, explicatif, etc.). Il définit une structure textuelle caractéristique, comme par exemple le schéma narratif : situation initiale - élément perturbateur - péripéties - dénouement - situation finale.

Le niveau de la *communication* correspond au contexte dans lequel le texte a été produit. Il détermine son objectif. La compréhension du texte par le lecteur, et donc la représentation mentale qu'il en construit, sera problématique si la structure du texte ne correspond pas à son type et/ou si le contexte de production du texte est en inadéquation avec l'objectif poursuivi. Ces éléments sont importants pour comprendre sur quoi reposent les ambiguïtés ou incompréhensions lors de la lecture d'un texte.

De ce fait, la compréhension d'un texte passe par la représentation mentale que s'en font les lecteurs. Les différents niveaux de celle-ci contribuent à la cohérence du texte dans son

ensemble. Cette représentation mentale cohérente est le « produit fini » (Rondelli, 2010a) du lecteur, ce qu'il a en tête à la fin de sa lecture. Le rédacteur doit donc veiller à anticiper la représentation mentale que va se construire le lecteur et de ce fait, réaliser les opérations dans l'ordre inverse. Cela conduit Chuy et Rondelli (2010 : 86) à les synthétiser, de manière relativement prescriptive, comme suit :

> « […] le scripteur doit créer un message clair expliquant pourquoi le lecteur a intérêt à lire son texte (niveau de *communication*) ; il doit disposer d'un répertoire étendu d'organisations textuelles pour sélectionner celle qui serait la plus adaptée à son intention de communication (niveau du *type de texte*) ; le rédacteur doit décider lesquelles de ces connaissances sont à inscrire explicitement dans le texte et lesquelles peuvent être omises pour être éventuellement inférées par le lecteur potentiel (niveau du *modèle de situation*) ; il doit pouvoir construire des réseaux de micro- et macro-propositions explicites qui permettront au lecteur de détecter les relations de cohérence locale et globale (niveau de *la base de texte*) ; finalement, il doit disposer d'un vocabulaire assez riche, des règles de grammaire et de syntaxe pour pouvoir encoder l'information (le niveau du *code de surface*). Si les exigences de ces cinq niveaux de représentation ne sont pas remplies, le texte final risque de ne pas montrer un niveau de cohérence reconnu suffisant. »

L'émission d'un message passe ainsi d'un modèle multidimensionnel présent au niveau cognitif (Fayol, 2002) à un message linéaire sous forme de discours (oral ou écrit). La cohérence est donc une construction du lecteur qui comporte une part de subjectivité (cf. jugement de cohérence, chapitre 4), ce qui explique la difficulté qu'on a à la définir.

Toutefois, d'un point de vue strictement linguistique, des chercheurs ont tenté de formaliser la notion de cohérence textuelle sous la forme de règles qui régissent la compréhension et la production d'un texte. Nous les présentons dans la section suivante.

1.1.2. L'approche linguistique : les quatre méta-règles de cohérence textuelle, selon Charolles (1978)

À l'échelle linguistique, déterminer les éléments en jeu dans la cohérence textuelle reste complexe, notamment à cause de l'aspect subjectif lié à l'interprétabilité des textes. Néanmoins, Charolles (1978), et après lui, Reinhart (1980), a tenté de dégager quelques « règles de bonne formation du texte ». À cette fin, il étudie les interventions des enseignants sur les énoncés malformés des élèves, énoncés jugés incohérents par les enseignants. De ces travaux, on retient que la cohérence textuelle relève de quatre règles principales (ou méta-

règles) : la méta-règle de *répétition*, la méta-règle de *progression*, la méta-règle de *non-contradiction* et la méta-règle de *relation*. Ces règles structurent à la fois les plans local (microstructurel) et global (macrostructurel) du texte.

La première de ces méta-règles est celle de **répétition.** Elle implique que « pour qu'un texte soit […] cohérent, il faut qu'il comporte dans son développement linéaire des éléments à récurrence stricte » (Charolles, 1978 : 14). Cette règle implique l'utilisation des procédés de rappel de certains éléments d'information[2]. Nous inventorions ces procédés ci-dessous en les illustrant par des exemples empruntés à l'auteur (les éléments en italiques indiquent les phénomènes de reprise) :

- la pronominalisation, en (1), dans lequel le SN indéfini expansé *une vieille dame* est repris par le pronom personnel *elle* dans la phrase qui suit :

  (1) *Une vieille femme* a été assassinée la semaine dernière à Besançon. *Elle* a été retrouvée étranglée dans sa baignoire.

- la définitivisation, en (2), dans lequel le SN indéfini complexe *une maison avec une grange* est repris par le SN défini *la maison* :

  (2) Jérôme vient d'acquérir *une maison avec une grange. La maison* est grande et a du cachet.

  et la référenciation déictique contextuelle, en (3), dans lequel le SN indéfini *une maison* est repris par le SN démonstratif, *cette maison* :

  (3) Jérôme vient d'acheter *une maison. Cette maison* est grande et a du cachet.

- la substitution lexicale, en (4), dans lequel le SN indéfini *un crime* est repris par un SN démonstratif contenant non pas comme dans les exemples précédents, la même tête lexicale, mais un hyponyme, *cet assassinat* :

  (4) Il y a eu *un crime* la semaine dernière à Besançon : une vieille dame a été étranglée dans sa baignoire. *Cet assassinat* est odieux.

- la reprise d'inférence, en (5). L'interprétation de l'exemple (5) nécessite un calcul interprétatif de la part du lecteur pour comprendre que si Nicole s'est inscrite à

[2] Ces procédés de rappel sont considérés comme relevant de la cohésion pour Halliday & Hasan (1976), mais nous en reparlerons plus loin (cf. ce chapitre, section 1.3.)

l'université, c'est qu'elle a obtenu son baccalauréat et a donc terminé ses études secondaires :

(5) Nicole s'est inscrite à l'université. Depuis qu'elle a fini ses études secondaires, elle ne sait plus vers quel métier s'orienter.

La méta-règle de répétition est reprise par Reinhart (1980 : 168) sous le nom de *connectedness* et énoncée comme suit : « A text is connected (cohesive) if each pair of its sentences […] is either : 1. Referentially link […], or 2. Linked by a semantic sentence connector »[3]. Toutefois, la condition de *connectedness* pour l'auteure se situe à un niveau légèrement différent de ce que Charolles (1978) regroupe sous sa condition de répétition. En effet, pour Reinhart, un texte est cohérent lorsque trois conditions sont réunies : *connectedness* (qu'elle appelle aussi *cohesion*), *consistency* et *relevance*. La condition de *connectedness* relève donc déjà de la cohésion pour Reinhart, alors que Charolles fait de son critère de répétition une condition de cohérence uniquement et différencie nettement cohérence et cohésion.

La méta-règle de **progression** implique que pour respecter la cohérence d'un texte, « il faut que son développement s'accompagne d'un apport sémantique constamment renouvelé » (Charolles, 1978 : 20), afin d'éviter un texte comme (6), emprunté à l'auteur :

(6) Le forgeron est vêtu d'un pantalon noir et d'un chapeau clair et d'une veste grise et marron foncé. Il tient à la main le piquet de la pioche et il tape dessus avec un marteau sur le piquet de la pioche. Les gestes qu'il fait, il a le piquet de la pioche et avec son marteau il tape sur le piquet de la pioche […].

Dans cet exemple issu d'un texte d'élève, le même objet de discours (*le piquet de la pioche*) est repris sans cesse sous la même forme et sans que soit apportée d'information nouvelle. Le lecteur peut avoir l'impression de ne pas avancer. Pour éviter de produire cet effet de piétinement, l'emploi de connecteurs et de différents types de progression thématique (cf. ce chapitre, 1.2.2.) est conseillé car « la production d'un texte cohérent suppose donc que soit réalisé un périlleux équilibre […] entre continuité thématique et progression sémantique (ou rhématique) » (*Ibidem : 21*).

Comme l'introduction de nouveaux référents ou de nouvelles informations ne peut être aléatoire, la méta-règle de **non-contradiction** (*condition of consistency* chez Reinhart, 1980) consiste en un « […] développement n'introduis[ant] aucun élément sémantique contredisant

[3] « Un texte est connecté (cohérent) si chaque couple de phrases […] est à la fois : 1. Lié référentiellement […], ou 2. Lié par une phrase faisant office de connecteur sémantique.» (notre traduction)

un contenu posé ou présupposé par une occurrence antérieure ou déductible de celle-ci par inférence » (Charolles, 1978 : 22). Autrement dit, un texte cohérent exclut la présence de phrases sémantiquement contradictoires. Notons au passage que ce genre de phrase est pourtant souvent présent dans les textes d'élèves. Il peut s'agir de contradictions énonciatives temporelles, comme dans (7), emprunté à l'auteur (et tous les autres exemples également) :

(7) Hier, il a mangé des gaufres au sucre. C'est sa mère qui les avait faites. Ensuite, il va dans le jardin pour jouer.

Dans cet exemple, il y a contradiction entre les deux premières phrases qui sont au passé et la troisième au présent, accentuée par la présence du connecteur *ensuite* qui devrait relier deux propositions dont les verbes sont présents dans le même système temporel.

Les contradictions peuvent également être discursives, comme dans (8)

(8) Casimir part en promenade avec monsieur du Snob. Arrivé à un arbre nous arrêtons pour parler. François et Julie veulent jouer à un jeu.

Ici, il y a contradiction au niveau de la narration : dans la première phrase, il est question de Casimir et d'un monsieur en narration à la troisième personne, puis on passe à une phrase avec la première personne du pluriel, *nous*, puis on change encore une fois avec l'apparition des personnages François et Julie.

Il peut s'agir aussi de contradictions inférentielles, comme dans (9) où la première phrase implique que Marie ne s'est pas mariée avec Paul, alors que la seconde phrase fait comprendre au lecteur qu'elle l'a épousé :

(9) Si Marie m'avait écouté, elle n'aurait pas définitivement repoussé Paul qui voulait l'épouser. Maintenant qu'elle est sa femme, elle s'en mord les doigts.

Enfin, il peut s'agir de contradictions de représentation du monde. Dans (10), rencontrer un arbre dans un couloir de métro est incohérent :

(10) Oscar sortit du métro. Il courait tête baissée dans un couloir quand il heurta de plein fouet un platane.

Enfin, la dernière méta-règle, qui est la méta-règle de **relation** (*relevance* chez Reinhart, 1980), stipule que « pour qu'une séquence ou qu'un texte soit cohérents, il faut que les faits qu'ils dénotent dans le monde représenté soient directement reliés » (Charolles,

1978 : 31). En effet, si les énoncés sont trop éloignés sémantiquement parlant, leur compréhension en est difficile voire impossible, comme dans l'exemple qui suit :

(11) Marie a acheté une armoire Louis XV. Elle a des migraines épouvantables.

Ici, le lecteur a du mal comprendre en quoi le fait d'acheter une vieille armoire est lié au mal de tête de Marie.

Ainsi, la compréhension - mais de fait, la production également - d'un message écrit répond à des règles linguistiques précises qui sont de l'ordre de la cohérence. Cependant, elles ne suffisent pas à elles seules à rendre un texte cohérent. En effet, l'encodeur doit aussi effectuer un travail de conservation/progression de l'information (Schneuwly, 1988), ce qui implique une analyse de l'organisation de l'information ou *statut informationnel*.

1.2. Le statut informationnel

Le statut informationnel, dans un discours, repose sur les différences de mise en œuvre de l'information nouvelle par rapport à celle qui est ancienne (ou connue) mais aussi sur le critère de la saillance, c'est-à-dire les facteurs qui rendent une information plus « proéminente » qu'une autre. Nous nous proposons dans cette section de définir ces notions.

1.2.1. L'information nouvelle *vs* l'information ancienne

Un discours étant une succession d'énoncés (ou d'informations) ayant un sens, pour assurer sa bonne interprétation, le scripteur doit maîtriser le statut de l'information. En effet, si nous revenons un instant à la règle de **progression** (cf. ce chapitre, 1.1.2), elle stipule qu'un texte doit apporter régulièrement de nouvelles informations. Pour illustrer cette règle, nous empruntons un exemple à Agatha Christie (12) :

(12) [M. Poirot] s'éveilla vers 8 heures et demie et se rendit au wagon-restaurant dans l'espoir d'une tasse de café brûlant. Il n'y trouva qu'une seule consommatrice, bien évidemment la jeune anglaise dont avait parlé le conducteur. Elle était grande, mince et brune, et on lui aurait donné dans les vingt-huit ans. (*Le crime de l'Orient-Express*, Agatha Christie : 14)

Dans cet exemple, on apprend les faits et gestes du personnage d'Hercule Poirot, avant d'en venir à l'introduction d'un personnage féminin, sur lequel on apprend aussi un certain nombre d'informations.

Non seulement ce texte renouvelle l'apport de nouvelle information (introduction d'Hercule Poirot et description de ce qu'il fait puis introduction de l'Anglaise et description de son apparence), et cette information est par la suite maintenue d'une phrase à l'autre et développée. Ainsi, un discours doit comporter deux types d'informations : nouvelle et ancienne. L'information nouvelle (ou *new contract* pour Clark & Haviland, 1977) désigne un référent (au sens le plus large : personnage, lieu, évènement, etc.) qui n'a pas encore été mentionné dans le discours et qui n'est pas accessible dans la situation d'énonciation par l'interlocuteur. L'ancienne information (ou *given contract*, *ibid.*) quant à elle, désigne un référent récupérable par l'interlocuteur dans le discours ou le contexte.

À ces deux niveaux d'information peuvent s'ajouter des niveaux intermédiaires, en une sorte de hiérarchisation dont le but est de « montrer que les procédés morphosyntaxiques constituent une sorte de codage des degrés de connaissances partagées » (Combettes, 1992). Aussi, Levelt (1989 : 99) propose quatre degrés « d'accessibilité » des référents, ici dans le sens de *givenness* en lien avec *given/new contract* de Chafe (1987) (à ne pas confondre avec la notion d'*accessibility* d'Ariel (1988), que nous développerons en 1.2.2) :

- Inaccessible : le référent est complètement nouveau. Ce degré correspond également à ce que Prince (1981) appelle *brand new* dans son échelle de « familiarité » (cf. ce chapitre, 1.2.2)
- Accessible hors du discours : le référent est inférable par le contexte ou les savoirs partagés d'une communauté
- Accessible dans le discours : le référent est déjà introduit précédemment dans le discours, donc récupérable facilement par l'interlocuteur
- *In focus* : le référent a été introduit préalablement et est activé dans le focus d'attention.

Le statut de l'information est important : selon que l'information est nouvelle ou ancienne, elle n'occupera pas la même place dans la structure syntaxique. En effet, un principe universel d'organisation des langues (Lambrecht, 1994) veut que l'information ancienne soit située préférentiellement en début de phrase, notamment en position de sujet syntaxique et très souvent sous la forme d'un pronom. L'information nouvelle serait, elle,

plutôt en fin de proposition, en position non sujet, car elle nécessite plus de ressources cognitives. L'information ancienne étant active, elle est rapidement accessible par l'interlocuteur, au contraire de l'information nouvelle. Il existe donc, dans un énoncé prototypique, une contrainte conversationnelle : le locuteur/scripteur ne doit pas transmettre plus d'une information nouvelle par phrase (*the one new idea constraint*, Chafe, 1994 : 119)

1.2.2. L'impact du statut informationnel sur le choix des formes linguistiques : l'accessibilité des référents

Pour introduire et maintenir une information/un référent, l'encodeur du message utilise des outils linguistiques et cohésifs. Selon le statut informationnel du référent, celui-ci est encodé sous différentes formes et distribué différemment dans la phrase : le statut informationnel est donc en lien avec les formes linguistiques choisies, comme le rappelle Mazur-Palandre (2009).

Dans cette optique, Prince (1981), et à sa suite Gundel & *al.* (1993, 2000) avec la hiérarchie du *donné*, proposent une tentative de classification des référents en fonction des divers degrés de connaissance partagée (ou *statuts cognitifs* chez Gundel & *al.)*. À cette fin, les référents sont classés du moins connu au plus connu, sur une échelle de familiarité (*givenness hierarchy*) que nous illustrons ici par les exemples de Combettes (1992) :

- Référents entièrement nouveaux (*brand-new*) : ils se trouvent au degré le plus bas de l'échelle, c'est-à-dire qu'ils ont la familiarité la plus faible. Ils sont désignés par des SN indéfinis, définis ou des noms propres suivis de détermination relativement complexes. Les référents entièrement nouveaux peuvent être rattachés au contexte/ancrés (par l'intermédiaire d'un modifieur), comme dans (13) où la phrase entre crochet présente la personne du vice-roi *d'Almeida* et la phrase qui suit présente *Albuquerque* comme *le successeur d'Almeida* :

  (13) [… le vice-roi, Francisco d'Almeida…] Le successeur d'Almeida, Albuquerque, est plus ambitieux.

Les référents entièrement nouveaux peuvent aussi ne pas être rattachés au contexte / non ancrés (leur existence est ignorée du décodeur) comme dans (14) avec le SN indéfini *un tunisien qui parlait l'espagnol* :

(14) Grâce à un tunisien qui parlait l'espagnol, Gama entre en relation avec le Zamorin, souverain de l'endroit.

- Référents inférables : il s'agit des référents qui peuvent être inférés à partir de connaissances conceptuelles, comme dans (15) où l'on sait que lorsqu'il est question d'un navire, il y a forcément des marins :

(15) […] quatre navires quittent le Tage le 8 juillet 1497. Le recrutement *des marins* a été difficile.

ou bien à partir de connaissances discursives comme dans (16) où l'on comprend que *certains* renvoie aux marins de la phrase précédente :

(16) Le recrutement des marins a été difficile. *Certains* ne croyaient-ils pas à l'existence de puissants aimants qui attiraient les navires ?

Les référents inférables sont souvent désignés par des SN définis ou des noms propres.

- Référents évoqués dans la situation : ils sont présents dans la conscience des locuteurs car déjà explicitement mentionnés dans le texte (17) ou présents dans la situation énonciative (18) (évidence perceptive) :

(17) Certains pensaient que […]. Or *ces idées* vont se modifier…
(18) Tu as lu *ce livre* ? (accompagné d'un geste ostensif)

Ce sont généralement des SN définis, des SN démonstratifs et des pronoms.

Pour résumer les différents degrés de l'échelle de familiarité de Prince, nous reprenons le schéma ci-dessous, emprunté à Gundel & *al.* (2000) et traduit par eux.

en focus >	activé >	familier >	uniquement identifiable >	référentiel >	identifiable quant au type
{*il*}	*cela, ceci, ce* N-*ci*, *celui-ci, celui-là*	{*ce* N-*là*}	{*le/la* N}	{ce N(-*ci*) *indéfini*}	{*un/une* N}

Figure 1 : Schéma de la *Givenness Hierachy*, échelle de familiarité (Prince, 1981, repris et traduit en français par Gundel & al., 2000)

L'échelle de familiarité ainsi décrite met en évidence le fait que les expressions linguistiques utilisées pour représenter un référent marquent son degré d'accessibilité (Givon, 1983 ; Ariel, 1988). Ariel (2001 : 29) résume sa théorie de l'accessibilité de la manière suivante :

> « Accessibility theory […] describes how human language, specifically, the referential system, is responsive to facts about human memory, where memory nodes are not equally activated at any given time. Some are highly activated, others are only mildly activated (…).[4] »

Dans un discours, le locuteur/scripteur adapte le choix des expressions référentielles à l'accessibilité supposée des référents visés. Cette accessibilité est déterminée en fonction de la quantité d'information fournie par l'expression, selon le principe de pertinence (Sperber & Wilson, 1989). Ainsi, moins accessible sera un référent, plus grand sera le contenu descriptif de l'expression choisie pour le désigner : « […] the lower the accessibility marker, the more lexical information it normally incorporates.[5] » (Ariel, 1988 : 82). Et inversement, plus un référent sera accessible, moins grand sera le contenu descriptif nécessaire pour son identification. De cette manière, les pronoms marquent une accessibilité forte (contenu descriptif faible), comme dans (20), alors que les SN définis ou les noms propres marquent une accessibilité plus faible (contenu descriptif fort), comme dans (19) :

(19) *Jean* est entré dans le magasin.

[4] « La théorie de l'accessibilité […] décrit comment le langage humain, de manière plus spécifique, le système référentiel, est réactif aux faits concernant la mémoire humaine, où les nœuds mémoriels ne sont pas activés de manière égale au même moment. Certains sont hautement activés, d'autres sont seulement moyennement activés (…). » (notre traduction)

[5] « […] moins un marqueur est accessible, plus grand sera le contenu lexical. » (notre traduction)

(20) *Il* est entré dans le magasin.

Si l'on veut être plus précis, il faut remarquer qu'Ariel (1988) propose dans sa théorie trois degrés d'accessibilité : *low accessibility*, *mid accessibility*, *high accessibility* et subdivise encore chacun en fonction des expressions linguistiques. Un des reproches fait à la théorie de l'accessibilité d'Ariel concerne sa trop grande flexibilité : « un marqueur référentiel peut, en fait, servir pour des référents d'accessibilités différentes » (Kleiber, 1990 : 250) (pour une critique de ce modèle et ceux qui ont à voir avec le statut cognitif des référents, voir Kleiber, 1990, 2001 ; Apothéloz, 1995a ; De Mulder, 2000 ; Cornish, 2014). Pour autant, ce modèle a le mérite de montrer que le locuteur/scripteur ne peut choisir la bonne expression référentielle sans prendre en compte le degré de connaissance partagée par l'interlocuteur/lecteur car « (…] s'il veut que son interlocuteur comprenne de qui ou de quoi il s'agit il a tout intérêt à utiliser une expression qui permette de trouver ou de retrouver le référent visé » (Kleiber, 1990 : 249).

Ainsi, le statut informationnel des référents (connu/donné ou nouveau) entre en jeu non seulement dans le choix des expressions référentielles, nous venons de le voir, mais il est aussi facteur de cohérence d'un texte. Par le changement de statut de l'information, c'est-à-dire le passage de l'information nouvelle à l'information ancienne (et vice-versa), il y a progression de l'énoncé et l'apport sémantique est constamment renouvelé. Cette progression, appelée *progression thématique*, contribue donc à la bonne compréhension d'un énoncé.

1.2.3. La progression thématique

Lorsque l'on se place d'un point de vue sémantico-pragmatique, on distingue dans un énoncé, deux constituants informationnels : le *thème* (ce dont on parle) et le *rhème* (ce qu'on en dit). Le thème peut correspondre au sujet grammatical, au niveau des référents (Lyons, 1980) et doit être en position initiale dans la proposition (Halliday, 1967). Le rhème correspond ordinairement au prédicat. Charolles (2002 : 153) résume ainsi :

> « […] un syntagme thématique [fait] allusion soit à une entité qui se trouve déjà au centre des préoccupations des participants à la communication, soit à une entité qu'ils n'auront aucun mal à rattacher à celles-ci. Le groupe verbal est destiné, par contre, à faire avancer l'information, autrement dit à supporter l'essentiel du contenu rhématique. »

Généralement, lorsque deux phrases se succèdent, le rhème de la première devient le thème de la seconde (Combettes, 1983). De cette manière, le rhème devient saillant dans la deuxième phrase. Il s'agit là d'une progression linéaire, comme dans (21), emprunté à l'auteur ainsi que les exemples suivants :

(21) *Le chat* court derrière *la souris*. *La souris* court derrière *un morceau de fromage*.

Combettes (1983) distingue également deux autres types de progression thématique : à thème constant et à thème dérivé. La progression à thème constant est caractérisée par le fait que le thème reste le même d'une phrase à l'autre malgré le changement de rhème :

(22) *Jean* sort de chez lui. *Il* caresse son chien au passage. Puis *il* monte en voiture et s'en va.

La progression à thème dérivé s'organise autour d'un hyperthème décliné en plusieurs thèmes (23) :

(23) *Plusieurs médailles* ont été distribuées aujourd'hui. *La médaille d'or* est revenue aux français. *La médaille d'argent* a été gagnée par les anglais.

Le thème est donc une information connue du locuteur, saillante parce qu'activée cognitivement (Combettes, 1977 ; Lyons, 1980). Le rhème est une information nouvelle. Ainsi, statut informationnel et progression thématique sont liés non seulement entre eux mais tous deux sont également associés à la notion de saillance.

1.2.4. La notion de saillance

Lors de l'interprétation d'un message, certains mécanismes linguistiques et cognitifs dirigent l'attention du sujet sur un élément plutôt qu'un autre et favorisent sa prise en compte dans le processus d'interprétation, au niveau de la détermination du sens comme de la résolution des références et coréférences (Landragin, 2004). La saillance est un élément prépondérant dans la production ou la compréhension d'un message car c'est un facteur de cohérence textuelle. Concept multicritère, la saillance repose sur des aspects prosodiques, lexicaux, morphologiques, syntaxiques, sémantiques, pragmatiques, stylistiques ou cognitifs (Schnedecker, 2009 ; Landragin, 2010). C'est donc elle qui rend un élément prégnant ou non, comme l'explique Landragin (2004 : 3) :

> « Du point de vue de la forme, être saillant, c'est ressortir en premier lors de l'audition d'un énoncé (ou de la lecture d'une phrase) compte tenu de la forme que prend cet énoncé, c'est-à-dire des mots qui le composent, de l'ordre et des fonctions grammaticales de ces mots. La saillance dénote alors le degré d'influence de l'attention du sujet par les entités perçues. Du point de vue du sens de l'énoncé, être saillant, c'est ressortir en premier lors de l'audition d'un énoncé, compte tenu du sens immédiatement perceptible de cet énoncé, c'est-à-dire avant tout processus complexe de raisonnement. »

La saillance permet donc de rendre un élément prégnant, que ce soit au niveau de sa forme (saillance linguistique formelle) ou de son fond (saillance linguistique sémantique).

Saillance linguistique formelle

La forme de l'énoncé est le point de départ pour décrire les différents facteurs de saillance, à l'oral comme à l'écrit, mais nous nous focaliserons ici sur les caractéristiques de la saillance dans le discours écrit, en nous appuyant principalement sur les travaux de Landragin (2004).

Dans l'aspect formel, lorsqu'on se place au niveau du mot, certains mots sont plus saillants que d'autres par leur nature même, comme c'est le cas avec les mots ayant une orthographe inhabituelle par exemple (mot avec une succession rare de consonnes), car ils sont remarquables au premier coup d'œil. Si l'on considère la classe grammaticale des mots, tous n'ont pas le même degré de saillance. C'est le cas notamment du nom propre : « A named character is more highly focused than one described by a referring expression[6] » (Garrod & Sanford, 1988).

Lorsque l'on se place au niveau de la phrase, la structure syntaxique est un facteur important de saillance. En effet, certaines constructions typiques de l'oral sont connues pour mettre en relief une entité que l'on veut rendre saillante. Ces constructions se retrouvent à l'écrit notamment chez les apprentis scripteurs. Il s'agit des constructions clivées en *c'est* (24) ou en *il y a* (25) mais aussi du détachement en tête de phrase (26) ou des dislocations (27) (Landragin, 2004). Les exemples suivants sont empruntés à l'auteur :

> (24) C'est le triangle rouge qui doit se mettre à côté du bleu.
>
> (25) Il y a le téléphone qui sonne.

[6] « Un personnage nommé [*ndlr.* par un nom propre] est plus focalisé qu'un personnage décrit par une expression référentielle » (notre traduction)

(26) Jean, sa sœur, je la déteste.

(27) Le triangle rouge, il doit se mettre à côté du bleu.

L'ordre et la fréquence des mots dans une phrase ont aussi un rôle. Généralement, les positions de début et de fin de phrase sont propices à rendre saillant le mot ou groupe de mots qui s'y trouve. De même, plus un mot est mentionné dans un énoncé (*via* une répétition, clarification, reprise, etc.) plus il est saillant (Landragin, 2004). Enfin, la fonction grammaticale d'un mot ou groupe de mots influe aussi sur son degré de saillance. Dans une langue à structure SVO (Sujet-Verbe-Objet) comme le français, le sujet, ordinairement en première position dans la phrase, est considéré comme le constituant grammatical le plus saillant. La Théorie du Centrage d'attention (Grosz & *al.*, 1995) suggère d'ailleurs qu'après le sujet, le complément d'objet est l'élément le plus saillant, suivi par les autres catégories grammaticales. Mais les seuls critères formels ne peuvent suffire pour déterminer la saillance d'une entité.

Saillance linguistique sémantique

D'un point de vue sémantique, les caractéristiques des entités représentées par les mots ont une influence sur leur saillance (Lyons, 1980 ; Pattabhiraman, 1993). En effet, le caractère « humain » ou « animé » d'une entité augmente sa saillance, par rapport à une entité « non animée » ou un animal, car comme l'explique Lyons (1980), il existe une hiérarchie universelle qui induit un intérêt décroissant selon les traits : « en tant qu'êtres humains nous nous intéressons davantage aux personnes qu'aux animaux, aux animaux qu'aux entités inanimées, et ainsi de suite. » (Lyons, 1980 : 142). Dans l'exemple (28), emprunté à Landragin (2004), *la foudre* est le sujet syntaxique et il est placé en première position, il devrait donc être le plus saillant (cf. *supra,* ordre et fréquence des mots). Or, *la foudre* étant une entité non humaine, elle entre en compétition avec l'entité humaine *la fillette*. La saillance se porte donc sur *la fillette* :

(28) La foudre a tué une fillette de huit ans : elle courait sur la plage au beau milieu de l'orage.

En outre, dans les cas où toutes les entités sont des humains, Wanner & Bateman (1990) avancent que les relations d'influence que les personnages humains exercent les uns sur les autres déterminent leur saillance, avant même la nature de l'évènement qui les lie. C'est le cas dans l'exemple de Landragin (2004) ci-après :

(29) L'élève écoute le professeur.

Dans (29), le professeur serait rendu saillant par le seul lien d'influence qu'il exerce sur son élève du fait de sa fonction.

Ainsi, la notion de saillance tout autant que celle de statut informationnel (et par extension, la progression thématique) sont des facteurs de cohérence textuelle car ils régissent le niveau macrotextuel, ou pour reprendre l'expression de Charolles (1988b), *l'interprétabilité du texte*. Mais comme nous l'avons vu plus haut (cf. ce chapitre, 1.1), la cohérence ne suffit à elle seule à faire qu'un texte soit perçu comme tel. Il faut aussi des marques linguistiques qui assurent les liens intra- et interphrastiques ou *cohésion*.

1.3. La cohésion textuelle selon Halliday & Hasan (1976)

Dans les travaux linguistiques sur la cohésion et les marques cohésives, les précurseurs sont sans conteste Halliday & Hasan (1976). Il leur est souvent reproché une description purement linguistique des marques de la cohésion, s'appuyant sur les opérations interprétatives uniquement *dans* le texte au détriment des opérations cognitives de *représentation mentale* (pour une critique de leurs travaux, voir entre autres Brown & Yule, 1983 ; Blass, 1990). Il est néanmoins indéniable que ces travaux fondateurs sur la cohésion textuelle apportent des éléments descriptifs. Halliday & Hasan (1976 : 6) définissent la cohésion ainsi : « The concept of cohesion is a semantic one; it refers to relations of meaning that exist within the text, and that define it as a text. »[7].

Selon ces chercheurs (*ibid.* : 10), « cohesion refers to the range of possibilities that exists for linking something with what has gone before »[8]. Ces liens entre entités (au sens le plus large du terme, que ce soient des personnages, des lieux ou même des évènements), ou « marques de relation entre énoncés » (Charolles, 1988b), sont des marques de cohésion participant à la cohérence générale du texte. Elles ont pour fonction « […] d'avertir le lecteur quand deux segments de texte doivent être connectés aux niveaux inter- et/ou intra-

[7] « Le concept de cohésion est un concept sémantique ; il fait référence aux relations de sens qui existent à l'intérieur du texte et qui le définissent en tant que texte » (notre traduction)

[8] « La cohésion fait référence à la gamme des possibilités qui existent pour relier quelque chose à un élément précédent. » (notre traduction)

phrastiques » (Chuy & Rondelli, 2010 : 88). Pour aider le lecteur dans cette entreprise, il existe différents types de marques de cohésion, dont nous allons faire un bref inventaire en nous appuyant sur les travaux et les exemples de Halliday & Hasan (1976). Toutefois, la différence de langue (anglais/français) pose parfois quelques limites à la transposition de ce modèle, ce que nous ne manquons pas de faire remarquer lorsque c'est le cas.

1.3.1. La substitution

La substitution est un « dispositif de cohésion lexico-grammatical » (Apothéloz, 1995a) car, au moins ancrée en partie dans la structure syntaxique, la substitution permet de maintenir une unité linguistique, par le remplacement d'un item par un autre, comme l'expliquent Halliday & Hasan (1976 : 89) : « Substitution, on the other hand, is a relation within the text. A substitute is a sort of counter which is used in place of the repetition of a particular item. »[9]

Les substitutions peuvent être :

- **nominale** : le substitut est le constituant tête du groupe nominal et remplace un nom ayant lui-même la fonction de constituant :

  (30) We have no coal fires ; only wood *ones*.
  « Nous n'avons pas de chauffage au charbon ; seulement au bois. » (notre traduction)

- **verbale** : Le principe est le même que pour la substitution nominale mais avec un verbe :

  (31) – Does Granny look after you every day ? – She can't *do* at weekends, because she has to go to her own house.
  « -Est-ce que mamie te garde tous les jours ? – Elle ne peut pas le faire les weekend, parce qu'elle doit rentrer chez elle. » (notre traduction)

[9] « La substitution, d'autre part, est une relation à l'intérieur du texte. Un substitut est une sorte de jeton qui peut être utilisé à la place de la répétition d'une entité particulière. » (notre traduction)

- **propositionnelle :** toute la proposition (sujet compris) est remplacée par le substitut :

(32) Everyone seems to think he's guilty. If *so,* no doubt he'll offer to resign.
« Il semble que tout le monde le pense coupable. S'il l'est, cela ne fait aucun doute qu'il proposera de démissionner. » (notre traduction)

La substitution est donc un outil qui permet de « faire figurer une information déjà donnée dans un environnement où est apportée de l'information nouvelle. » (Apothéloz, 1995a : 121).

1.3.2. L'ellipse

L'ellipse est un phénomène proche de la substitution par son fonctionnement. En effet, l'ellipse se manifeste par l'effacement d'une partie de la structure grammaticale d'une phrase sans en empêcher la compréhension, comme dans l'exemple ci-dessous :

(33) Here are my two white silk scarves. I can lend you *one* if you like.

« Voici mes deux écharpes blanches en soie. Je peux t'en prêter une si tu veux. » (notre traduction)

Ici, le *one* de la deuxième phrase remplace la reprise telle quelle de *white silk scarf* de la phrase précédente (il n'est pas écrit *I can lend you one white silk scarf if you want*). Les ellipses peuvent être nominales, comme dans l'exemple précédent, mais aussi verbales ou propositionnelles. Dans le cas de l'ellipse verbale, il s'agit d'effacer le verbe lexical, comme dans (34) où les opérateurs verbaux *might, was to, may not, should,* sont susceptibles d'être complétés par le verbe lexical :

(34) – Is John going to come ?
- He *might*. He *was to*, but he *may not*.
- He *should*, if he wants his name to be considered.

Cet exemple est difficilement traduisible en français car sa structure est typique de l'anglais mais nous tentons tout de même d'en donner une traduction : « - John va-t-il venir ? – Cela se pourrait. Il en avait l'intention, mais peut-être qu'il ne viendra pas. - Il devrait, s'il veut que son nom soit pris en compte. »

Mais l'ellipse verbale peut aussi consister en l'effacement de l'opérateur verbal, il s'agit alors du mécanisme inverse :

(35) –What will they be doing now, do you think ?

- *Finishing* their essays, probably.

  « Que sont-ils en train de faire, à ton avis ? – Terminer leur rédaction, probablement. » (notre traduction)

Dans le cas d'une ellipse propositionnelle, l'effacement se répercute soit sur l'élément modalisateur (le sujet grammatical de la proposition et la partie fléchie du groupe verbal), comme l'illustre (36) :

(36) - What was the Duke going to do ? – Plant a row of poplars in the park.

« Qu'est-ce que le duc était sur le point de faire ? – Planter une rangée de peupliers dans le parc. » (notre traduction)

soit sur l'élément propositionnel (la partie restante du groupe verbal et les éventuels compléments et adjoints) comme dans (37) :

(37) – Who was going to plant a row of poplars in the park? – The Duke was.

« -Qui était sur le point de planter une rangée de peuplier dans le parc ? – Le Duc. » (notre traduction)

L'ellipse a cependant besoin de s'appuyer sur une phrase précédente pour être pleinement comprise par le lecteur et de fait, ne fonctionne qu'à de très courtes distances.

1.3.3. La conjonction

Si l'ellipse et la substitution sont des outils cohésifs qui fonctionnent en renvoyant à de l'information en amont ou en aval du texte, le dispositif de *conjonction* caractérise la manière dont l'énoncé qui suit doit être relié à celui qui précède. Il n'y a donc plus de lien avec la structure syntaxique propre à la phrase, comme en témoigne l'exemple suivant :

(38) They fought a battle. Afterwards, it snowed.

« Ils menèrent un combat. Après cela, il neigea. » (notre traduction)

Apothéloz (1995a) explique que dans (38), il y a une succession de deux phrases, c'est-à-dire deux entités non reliées syntaxiquement. La relation de succession temporelle est indiquée par l'adverbe *après,* fonctionnant comme adjoint.

Lorsqu'il s'agit de la relation temporelle entre deux évènements narrés et que la temporalité est celle des évènements, la relation conjonctive est qualifiée d'*externe,* comme c'est le cas dans (39) où les deux situations se succèdent à l'intérieur de la narration :

(39) First he switched on the light. Next he inserted the key into the lock.

« D'abord, il a allumé la lumière. Ensuite, il a inséré la clé dans la serrure. » (notre traduction)

Lorsqu'il s'agit de la relation entre deux énoncés et que la temporalité est celle de la production de ces énoncés, la relation conjonctive est dite *interne*. C'est le cas dans l'exemple suivant, où les deux situations sont reliées par l'organisation du discours même (*first/next*), par la matérialisation des étapes de l'argumentation :

(40) First he was unable to stand upright. Next, he was incapable of inserting the key into the lock.

« D'abord il fut incapable de se tenir debout. Ensuite, il fut incapable d'insérer la clé dans la serrure. » (notre traduction)

1.3.4. La cohésion lexicale

Le seul dispositif cohésif défini par Halliday & Hasan (1976) qui ne présente pas de caractère grammatical est la *cohésion lexicale*. En effet, ce type de cohésion se manifeste dans le lexique, au niveau des choix lexicaux, sous deux formes différentes.

La première forme de cohésion lexicale, la *réitération*, se traduit par la répétition d'un lexème par le même lexème (41) ou bien à travers l'emploi d'un synonyme ou quasi-synonyme (42), d'un hyperonyme (43) ou d'un nom général (44) :

(41) There was a large *mushroom* growing near her, about the same height as herself ; […] She stretched herself up on tiptoe, and peeped over the edge of the *mushroom*.

« Il y avait un gros champignon qui poussait à côté d'elle, à peu près de la même taille qu'elle ; […] Elle s'étira sur la pointe des pieds et jeta un coup d'œil par-dessus le bord du champignon. » (notre traduction)

(42) Accordingly […] I took leave, and turned to the *ascent* of the peak. The *climb* is perfectly easy…

« Par conséquent […] je pris congé, et me tournai vers l'ascension du pic. L'escalade est parfaitement facile.» (notre traduction).

(43) Henry's bought himself a new *Jaguar*. He practically lives in the *car*.

« Henry s'est acheté une nouvelle Jaguar. Il vit pratiquement dans la voiture. » (notre traduction)

(44) Can you tell me where to stay in *Geneva* ? I've never been to the *place*.

« Pouvez-vous me dire où je peux séjourner à Genève ? Je ne suis jamais allé dans cet endroit. » (notre traduction)

La seconde forme de cohésion lexicale, la *collocation*, se manifeste lorsque des lexèmes ont tendance à se côtoyer dans les textes. Comme l'explique Alonso Ramos (1994), c'est à J.R. Firth que l'on doit le terme de *collocation* pour décrire des combinaisons récurrentes d'unités lexicales. Pour Firth, la collocation est la cooccurrence de deux éléments linguistiques, et l'attente mutuelle (*collocability*) que chaque mot fait porter sur l'autre. Par exemple, les lexèmes *dark* et *night* ont tendance à apparaître ensemble, comme l'explique Firth (1957 : 196) :

> « Meaning by collocation is an abstraction at the syntagmatic level and is not directly concerned with the conceptual or idea approach to the meaning of words. One of the meanings of *night* is its collocability with *dark*, and of *dark*, of course, collocation with *night.*[10] »

La collocation se traduit donc par certains types de relations entre deux termes : antonymie (*like/hate*[11]), complémentarité (*stand up/sit down*), converses (*buy/sell*), partie-tout (*car/brake*), etc.

[10] « Une collocation est une abstraction au niveau syntagmatique et ne se préoccupe pas directement du concept ou de l'approche du sens des mots. Un des sens de *nuit* est sa collocativité avec *noire*, et [un des sens] de *noire*, bien entendu, sa collocation avec *nuit*. » (notre traduction)

[11] Aimer/détester ; assis/debout ; acheter/vendre ; voiture/frein.

Tous les éléments que nous venons de voir participent à la cohésion d'un texte, à des niveaux différents. Les marques de cohésion contribuent donc grandement à la construction de la référence.

1.4. La continuité référentielle

Pour comprendre ce qu'est la référence, nous partirons de la question suivante : dans la vie quotidienne, comment parvenons-nous à parler de quelqu'un ou de quelque chose à notre interlocuteur afin qu'il comprenne de quoi ou de qui il s'agit ? En d'autres termes, comment parvenons-nous à référer au monde par l'intermédiaire du langage ? (Charolles, 2002). C'est à cette dernière question que nous allons tenter de répondre à présent, en décrivant d'abord succinctement la notion de référence (1.4.1.) puis en présentant les expressions linguistiques qui marquent la référence et la continuité référentielle (1.4.2.).

1.4.1. La description et le traitement des expressions référentielles

Comme les notions de cohérence et de cohésion, celle de référence a fait l'objet de très nombreux travaux. Nous ne prétendons pas ici à une revue exhaustive, ni à un tour complet de la question, mais présentons simplement quelques notions principales.

La référence est le rapport qu'entretient le langage avec ce qui lui est extérieur (Apothéloz & Pekarek Doehler, 2003). Dans la synthèse que font ces chercheurs sur les travaux concernant la référence, il apparaît bien que les perspectives d'analyse ont évolué à travers le temps. En effet, la référence a été longtemps étudiée dans une perspective textualiste dans laquelle la segmentalité et la séquentialité textuelles, éléments uniquement observés, ont mené à la primauté de la notion d'antécédent (nous reviendrons en détail sur cette notion dans le chapitre 2). Puis les mécanismes informationnels, mémoriels et inférentiels ont été pris en compte dans l'analyse, montrant ainsi l'importance de la représentation mentale et du savoir partagé dans la construction de la référence (entre autres, Chafe, 1987 ; Cornish, 1987 ; Berrendonner & Reichler-Béguelin, 1989 ; Givón, 1992, 1995b). Malgré ce changement de perspective dans l'étude de la référence, l'analyse de cette

notion porte sur des données monologales, écrites et hors de l'activité de discours dans sa dimension situationnelle, sociale et interactionnelle. Après cela, un nouveau courant d'étude de la référence a vu le jour et « cherche à intégrer les dimensions situationnelle, sociale et praxéologique des activités langagières » (Apothéloz & Pekarek Doehler, 2003 : 112). Au vu de notre sujet d'étude, nous nous inscrivons dans une approche « représentationaliste » de la référence, dans laquelle les processus référentiels s'appuient à la fois sur l'occurrence ou la cooccurrence des segments linguistiques dans un texte et sur les structures sémantiques de ce dernier, mais aussi sur le savoir partagé des interlocuteurs (ou *mémoire discursive,* Reichler-Béguelin, 1988a ; 1989 ; cf. ce chapitre, section 1.1.1.). Les processus référentiels jouent un rôle important dans la construction de la représentation mentale des interlocuteurs, qui elle-même participe à la construction de la référence. Nous reviendrons plus en détail sur ces notions dans le chapitre 2.

Deux éléments linguistiques évoquant la même entité entretiennent un lien de référence. Ainsi, dans l'exemple ci-dessous emprunté à Mazur-Palandre (2009), le pronom personnel *ils* réfère au syntagme nominal lexical *des enfants*, référent introduit précédemment :

(45) Il y a *des enfants* qui jouent dans la cour de récréation.
Ils jouent au chat et à la souris.

Halliday & Hasan (1976) ont décrit trois types de référence. Le premier, la *référence personnelle*, regroupe les éléments linguistiques qui marquent la personne (pronoms personnels, possessifs et déterminants possessifs). Le deuxième, la *référence démonstrative*, renvoie aux marqueurs démonstratifs (pronoms ou déterminants démonstratifs). Le troisième, la *référence comparative*, se rapporte aux éléments linguistiques concernant un sémantisme de comparaison (adjectifs et adverbes de comparaison). Dans la mesure où notre étude est fondée sur l'analyse de la référence dans un texte narratif et notre analyse portant sur des expressions linguistiques qui marquent la référence personnelle et la référence démonstrative, nous nous focaliserons sur ces deux types de référence et laisserons de côté le troisième type, en nous appuyant dans cette tâche sur le chapitre d'Apothéloz (1995a) qui présente les travaux d'Halliday et Hasan (1976) et reprend leurs exemples en anglais.

1.4.1.1. La référence personnelle

La référence personnelle regroupe à la fois les pronoms personnels, les déterminants possessifs et les pronoms possessifs. Il convient toutefois de distinguer les éléments de première et deuxième personne d'une part, et la troisième personne d'autre part. En effet, mis à part dans le discours direct, comme dans l'exemple (46), les deux premières personnes ne peuvent généralement être interprétées qu'en situation car il n'y a pas de référent accessible dans le co-texte. On est alors dans une situation *exophorique* (cf. chapitre 2 pour plus de détails).

(46) There was a brief note from Suzan. She just said, « I am not coming home this weekend »

« Il y avait un mot bref de Suzan. Elle disait seulement, « Je ne rentre pas ce week-end. » (notre traduction)

Ainsi, à l'écrit, seule la troisième personne est directement interprétable grâce au co-texte. L'interprétation dépend donc de la situation. On est alors en situation *endophorique* et c'est celle à laquelle nous allons le plus nous intéresser (cf. chapitre 2 pour plus de détails). Cette troisième personne, *il/elle* pour les pronoms personnels en français ou *he/she* en anglais, trouve le plus couramment son référent dans une expression nominale. Il en va ainsi avec (46) où *she* s'interprète comme faisant référence au personnage de *Suzan.* Cependant, lorsqu'il s'agit de la troisième personne générique, *it* en anglais (fréquemment traduit par *cela/ce/il* en français), il est nécessaire de recourir à un segment de texte entier et non plus une expression nominale pour pouvoir l'interpréter, comme dans l'exemple (47).

(47) [The Queen said :] « Curtsey while you're thinking what to say. *It* saves time. » Alice wondered a little at this, but she was too much in awe the Queen to disbelieve *it.*

« [La reine dit :] « Faites donc la révérence, pendant que vous réfléchissez à ce que vous allez dire. Cela fait gagner du temps. » Alice fut quelque peu étonnée d'entendre de telles paroles, mais elle était bien trop impressionnée par la Reine pour ne pas croire ce que celle-ci venait de dire. » (Traduction d'Apothéloz, 1995a).

Dans ce passage, le premier *it* ne peut se comprendre qu'en regard de la phrase précédente : il reprend le fait qu'Alice doit penser à ce qu'elle doit dire tout en faisant la révérence. Il s'agit d'un procès complexe. Dans ce cas, on parle de *référence étendue (extended reference)*. Le second *it* quant à lui, se comprend en prenant en considération le fait que cela fait gagner du

temps de faire la révérence pendant qu'elle réfléchit à ce qu'elle doit dire. Il s'agit d'un fait. On parle de *référence textuelle (textual reference).*

1.4.1.2. La référence démonstrative

Généralement, comme le précise Apothéloz (1995a), la référence démonstrative se présente sous la forme de démonstratifs adverbiaux (*here, there, now, then*/*ici, là-bas, maintenant, ensuite*) et de démonstratifs nominaux (*this, these, that, those/ ce-cette-cet, ces, ça).* En plus de ces deux catégories, Halliday et Hasan ajoutent le déterminant défini *the* (*le-la).*

Dans leur emploi endophorique, les déterminants démonstratifs « réfèrent à un ensemble plus large d'individus que l'expression ou les expressions nominales qui leur servent d'interprétant » (Apothéloz, 1995a :112), comme dans (48) où le démonstratif réfère à la fusion mentionnée dans le premier énoncé mais de manière plus large, à toutes les fusions en général.

(48) There's been another big industrial merger. It seems that nothing can be done about this.

« Il y a eu une autre grande fusion industrielle. Il semble que rien ne peut être fait contre cela. » (notre traduction)

Le déterminant défini *the* (*le*/*la*/*les* en français) signale que l'expression nominale qu'il détermine réfère à un objet singulier et identifiable. Mais ce déterminant est en lui-même dépourvu de tout contenu, à l'inverse des autres déterminants qui contiennent en eux-mêmes une certaine information référentielle (le possesseur pour les possessifs par exemple) (Apothéloz, 1995a). Ainsi, lorsqu'il est associé à un nom, le déterminant défini peut référer de manière exophorique ou endophorique. Dans ce second cas, le déterminant *the* peut être anaphorique, c'est-à-dire interprétable grâce à une expression qui se trouve avant lui, comme dans (49), ou cataphorique, c'est-à-dire qu'il est interprétable grâce à une expression linguistique qui se trouve après lui dans l'énoncé, comme dans (50) où par exemple, il ne s'agit pas de n'importe quelle ascension mais de celle du Mont Everest :

(49) She found herself in a long, low hall which was lit up by a row of lamps hanging from the roof. There were doors all round *the hall*, but they were locked.

« Elle se retrouva à présent dans une longue salle basse éclairée par une rangée de lampes accrochées au plafond. Il y avait plusieurs portes autour de la salle, mais elles étaient fermées à clé. » (notre traduction)

(50) *The ascent* of Mount Everest

The party in power

The best way to achieve stability.

« L'ascension du Mont Everest / Le parti politique au pouvoir / Le meilleur moyen d'atteindre la stabilité » (notre traduction)

Ainsi, dans son emploi endophorique, le déterminant *the* signale non seulement qu'il réfère à un objet identifiable, mais aussi que toutes les informations nécessaires à son identification sont disponibles dans le reste de l'énoncé, notamment par la contrainte syntaxique du fait de la présence du complément du nom.

Dans le cadre de notre étude, la référence endophorique nous intéresse particulièrement, notamment toutes ses manifestations sous la forme d'anaphores. Selon Mazur-Palandre (2009), pour produire un texte cohérent et cohésif, l'encodeur doit effectuer un travail de conservation/progression (Schneuwly, 1988) des syntagmes nominaux, ce qui implique une analyse de l'anaphore (introduction de nouveaux référents, reprise de ces référents, différenciation de ces référents), sur laquelle nous reviendrons en détail dans le chapitre 2, et une analyse de l'organisation de l'information (statut informationnel), que nous avons décrite plus haut (cf. ce chapitre, 1.2). Ainsi, la référence concourt à donner à un texte cohérence et cohésion par la répétition de certaines entités : c'est la *continuité référentielle* dont parle Apothéloz (1995a : 9-10) :

« Qu'est-ce qu'un texte ? […] Quelle différence entre un enchaînement de phrases dans lequel tout le monde s'accorde à voir un texte et un enchaînement de phrases dans lequel tout le monde s'accorde à ne voir qu'un enchaînement de phrases ? […] Or, on reconnaît généralement qu'un des dispositifs de cette macro-syntagmatique réside dans la répétition, partielle ou complète, modifiée ou non modifiée, de certains des objets (*référents*) qui sont désignés dans le texte. Pour qu'une suite de phrases puisse être perçue comme un texte, il faudrait, en d'autres termes, qu'elle présente une *continuité référentielle* minimale. Cette idée est généralement exprimée en disant que la continuité référentielle contribue à assurer au texte sa cohésion. »

La continuité référentielle a donc un lien direct avec la méta-règle de répétition de Charolles (1978 ; cf. ce chapitre, section 1.1.) puisqu'elle permet de maintenir un référent tout au long du discours.

1.4.2. Quelles formes linguistiques pour aider à construire la référence ?

L'acte de référence est le fait d'attirer l'attention d'un ou plusieurs interlocuteurs sur une ou plusieurs entités du monde qui les entoure (Charolles, 2002). Pour cela, la langue met à disposition des locuteurs plusieurs expressions linguistiques capables de marquer la continuité référentielle : les noms propres, les syntagmes nominaux SN et les pronoms. Nous nous attacherons ici à résumer les principales caractéristiques de ces formes de manière très synthétique, en nous appuyant principalement sur les travaux de Charolles (2002) et Schnedecker (1997).

1.4.2.1. Les noms propres

Comme le montrent les travaux de Schnedecker (1997), sur lesquels nous nous appuyons dans cette section, le nom propre est une des principales expressions linguistiques qui permet d'attirer l'attention de l'interlocuteur sur une entité dont le locuteur veut parler. Il se distingue du nom commun par le fait qu'il s'emploie sans déterminant, ne marque ni le genre ni le nombre et prend une majuscule à l'écrit. De ce fait, les noms propres peuvent référer à différentes sortes d'entités particulières : personnes, villes, fleuves, animaux… car « l'usage d'un nom propre ne signale aucune autre intention chez le locuteur que de viser un être unique » (Charolles, 2002 : 54).

En effet, le nom propre se différencie de la description définie en ceci qu'il désigne une entité particulière (*Jacques Dupont)* sous toutes ses formes et dans tous ses états, là où la description définie désigne une entité par des traits caractéristiques qui peuvent évoluer (Jacques Dupont, *l'actuel maire de la ville*). Quoi qu'il en soit, le nom propre est uni à son porteur par une relation à caractère dénominatif (Kleiber, 1981) : le nom propre *Untel* dénote

tout et rien que le particulier qui satisfait au prédicat de dénomination *s'appeler Untel* (Charolles, 2002).

Ainsi, le nom propre réfère à un particulier en le discriminant au sein d'une catégorie d'autres particuliers. Ce besoin de différenciation implique que l'entité à laquelle réfère le nom propre n'était pas encore au centre de l'attention de l'interlocuteur. Par cette opération, le nom propre porte l'attention sur une entité non encore saillante et signale qu'elle a de fortes chances de le devenir par la suite : « les noms propres réfèrent à des particuliers qui ne sont pas supposés déjà présents dans l'esprit des interlocuteurs : ils attirent l'attention sur ces particuliers en vue de renvois ultérieurs » (Charolles, 2002 : 74). En ce sens, ils sont la forme linguistique privilégiée lorsque l'accessibilité du référent est faible (théorie de l'accessibilité d'Ariel, cf. *supra* 1.2.). Autrement dit, les noms propres sont employés principalement en première mention du référent, comme pour l'introduire dans la sphère interlocutoire, ou bien le réinstancier lorsque le discours s'est trop éloigné du sujet premier et qu'il serait trop coûteux - cognitivement parlant - pour l'interlocuteur de comprendre que l'on parle à nouveau de l'entité désignée par le nom propre. En outre, le nom propre peut servir à hiérarchiser les référents parmi toutes les expressions référentielles d'un discours (l'entité qui est désignée à l'aide d'un nom propre sera plus importante que les autres).

1.4.2.2. Les déterminants des syntagmes nominaux

Les syntagmes nominaux (ou SN) sont sans doute les premières formes utilisées pour référer à une entité, qu'elle soit humaine, non humaine, animée ou non. Les noms peuvent être massifs (ils dénotent une entité non bornée, non dénombrable comme l'eau par exemple, ou le sable), ou comptables (ils dénotent une entité qui peut être bornée, dénombrable). Par ailleurs, un SN peut être générique, auquel cas il regroupe une classe entière d'individus ou d'objets, comme dans (51), ou bien spécifique (52), auquel cas il marque la référence à une entité unique (les exemples sont empruntés à Charolles, 2002) :

(51) Les chats sont très malins.

(52) Le chat a miaulé toute la nuit.

Si le nom fournit une indication sur la nature du référent, il ne faut pas oublier qu'il est généralement inséparable de son déterminant, dont le rôle est de coder des procédures

interprétatives subtiles (Charolles, 2002), dépendant de la classe grammaticale à laquelle il appartient : déterminants indéfini, défini, démonstratif. Le déterminant indéfini permet de fixer le référent et non de le désigner (comme c'est le cas des déterminants définis et démonstratifs), à l'aide de diverses opérations (instanciation d'une classe extensionnelle correspondant au descripteur nominal utilisé, extraction d'un ressortissant quelconque, attribution de propriété distinctive à l'aide de la prédication verbale, Charolles, 2002). Ainsi, l'énoncé entier participe à l'établissement de la référence et pas uniquement le seul SN indéfini.

Les SN indéfinis

Les SN indéfinis sont des expressions référentielles composées d'un déterminant indéfini *un/une* suivi d'un nom commun. Les expressions nominales indéfinies sont employées lorsque le locuteur introduit pour la première fois l'objet du discours, ou un nouveau référent (*brand new*, Prince, 1981). Comme le montre Charolles (2002 : 144), cela est possible car les SN indéfinis « ne sollicitent aucune préconception de l'entité visée et ne font appel à aucune connivence intellectuelle autre que celle relativement triviale, consistant à exploiter les ressources dénotationnelles des N et des prédicats ». Autrement dit, les SN indéfinis sont autonomes référentiellement parlant car ils ne demandent pas à ce que l'interlocuteur sache déjà de quoi il va être question dans l'énoncé.

Les SN définis

Les SN définis sont des expressions référentielles composées d'un déterminant défini *le/la/les* suivi d'un nom commun. Comme ce type de SN véhicule des informations sur les entités qu'il sert à désigner, on parle de *description.* Notons que le terme de *description* est valable aussi pour les autres types de SN : descriptions démonstratives, etc. Ces descriptions peuvent être complètes : le SN défini comporte des adjoints qui spécifient l'entité dont il est question (53), ou incomplètes (54). La différence entre les deux tient à leur autonomie référentielle : les descriptions définies complètes ne codent qu'une seule entité ou un seul groupe d'entités et permettent ainsi de les identifier indépendamment de toute information sur le contexte où elles sont employées (Charolles, 2002 ; les exemples suivants lui sont empruntés).

(53) *L'auteur de Notre Dame de Paris* est mort depuis bien longtemps.

(54) *L'auteur* est mort depuis bien longtemps.

Tout acte de référence accompli au moyen d'une expression définie « implique le repérage d'un ensemble d'objets à l'intérieur duquel le référent peut être identifié non ambigument » (Apothéloz, 1995a). Cette identification est rendue possible grâce à une description complète. Dans le cadre d'une description incomplète, l'identification se fait à l'aide du contexte restreint (appelé aussi *ensemble partagé*, Hawkins, 1978), *ensemble relationnel* (Galmiche, 1986), *domaine d'interprétation* (Corblin, 1987c) ou *circonstance d'évaluation* (Kleiber, 1992) et grâce notamment aux indices de saillance (cf. ce chapitre, 1.2). Lorsqu'il s'agit d'un SN défini incomplet, il doit y avoir dans le contexte un article indéfini avec lequel l'article défini entretient une relation anaphorique, comme l'explique Van de Velde (1997 : 84) : « tout article défini est appuyé sur un article indéfini avec lequel il entretient une relation anaphorique ». En effet, pour que l'interlocuteur réussisse à identifier sans problème le référent de la description définie incomplète, il faut préalablement attirer son attention sur une entité qui n'est pas encore focale dans son esprit (Charolles, 2002) : c'est justement ce que permet le déterminant indéfini, comme nous le verrons plus avant dans cette section.

Ainsi, le SN défini est en lien avec la représentation mentale que se font les participants lors d'un échange et c'est une forme qui « indique au destinataire que le référent appartient à un ensemble délimitable d'objets et que, pour identifier cet objet, il faut conjointement localiser cet ensemble et, dans cet ensemble, identifier le référent visé. L'identification proprement dite se fait sur la base des informations apportées par le constituant lexical de l'expression » (Hawkins, 1977a, cité par Charolles, 2002). Le déterminant défini sert donc à référer à un objet singulier (55), à une classe d'objets (56) ou à un objet « massif » (57), comme le montrent les exemples suivants, empruntés à Charolles (2002) :

(55) Elle attrapa *le ruban bleu*.

(56) *La vaisselle* n'est plus l'apanage des listes de mariage.

(57) *L'eau* est bonne.

Les SN démonstratifs

Les SN démonstratifs sont des expressions référentielles composées d'un déterminant démonstratif *ce/cet/cette/ces* suivi d'un nom commun. Ce sont des expressions déictiques ou anaphoriques qui nécessitent le recours à la situation d'énonciation pour être interprétées (Charolles, 2002). En effet, le démonstratif dirige l'attention sur une entité identifiable car immédiatement perceptible dans la situation d'énonciation (geste ostensif à l'oral par exemple) ou venant d'être évoquée dans le discours au moyen d'expressions linguistiques à l'écrit (contrainte d'unification ou *matching constraint* d'Hawkins, 1978), comme dans l'exemple suivant tiré de Charolles (2002) :

(58) A qui est ce parapluie ?

Les démonstratifs exploitent donc en premier lieu les caractéristiques du contexte matériel dans lequel ils sont employés. À l'inverse des SN définis qui identifient leur objet essentiellement en vertu de la relation lexicale qu'ils entretiennent avec le nom tête du SN (Corblin, 1987c, les qualifie alors d'*identifieur*), les SN démonstratifs peuvent être qualifiés de *reclassifieur* (Corblin, 1987c) car ils repèrent un objet en fonction de sa position de proximité (Charolles, 2002). Selon Marandin (1986 : 77), cette proximité induit la présence d'un antécédent avec lequel ils pourraient être mis en relation car « l'interprétation anaphorique d'un SN démonstratif […] requiert l'occurrence effective dans le contexte gauche d'un terme avec lequel [il peut] être mis en relation […] ».

En outre, De Mulder (1997), cité par Charolles (2002 : 120), explique que « contrairement au défini qui appréhende le référent comme unique du fait de la relation qu'il doit entretenir avec un des participants de la scène, le démonstratif implique une rupture avec cette scène ». De cette manière, le démonstratif situe le référent dans le focus d'attention de l'interlocuteur alors qu'il ne l'était pas auparavant, présentant ainsi l'entité comme étant nouvelle, ou la présentant d'un point de vue différent, comme l'explique Charolles (2002 : 121) : « le SN démonstratif installe le référent dans le focus d'attention des destinataires comme une entité nouvelle ou, en tout cas, en faisant abstraction de la représentation que les destinataires peuvent ou pourraient en avoir dans les circonstances où intervient l'échange ».

1.4.2.3. Les pronoms personnels

En français, la classe des pronoms regroupe des pronoms de divers types : personnels, démonstratifs, possessifs, indéfinis, interrogatifs, relatifs. Bien que chacun puisse mériter une étude détaillée, nous nous intéressons particulièrement à la classe des pronoms personnels, parce que ce sont les plus représentés dans les textes narratifs d'apprentis scripteurs.

Contrairement aux SN définis et démonstratifs que nous avons décrits *supra*, les pronoms personnels ne comportent pas de tête lexicale. Ils ne portent donc pas en eux-mêmes d'indication sur la catégorie de leur référent, mais cela ne les empêche pas de renvoyer à des particuliers et donc de référer de manière définie (Charolles, 2002). Comme l'expliquent Riegel & *al.* (1994), les pronoms sont des termes grammaticaux dont la morphologie change avec leur fonction grammaticale, mais aussi leur rang (la « personne »), nombre, genre et degré d'intégration dans la construction (formes conjointes ou clitiques et formes disjointes). Ainsi, il est important de distinguer les pronoms de rang 3 et 6, c'est-à-dire les pronoms personnels de troisième personne du singulier (3) et du pluriel (6), des pronoms de rang 1, 2, 4 et 5, c'est-à-dire les pronoms personnels de première et deuxième personnes du singulier et du pluriel[12], dits aussi *pronoms de dialogue*.

L'anaphorique « troisième personne »

Le pronom de troisième personne (rang 3 et 6 dans la conjugaison) n'est pas réservé aux humains. Mise à part la forme personnelle sujet indéfinie *on* qui n'est pas marquée en genre mais est réservée à des humains, le pronom de troisième personne peut référer à des entités non humaines (un animal par exemple) ou non animées (une table). Bien qu'il ne comporte pas de tête lexicale, comme il est marqué en genre (il est le seul, les pronoms de dialogue ne le sont pas), son emploi engendre la présence d'un antécédent : « la production et la compréhension d'un pronom marqué en genre suppose en effet que le locuteur et l'interprétant ont non seulement accès à la catégorie sous laquelle est conceptualisé le référent mais au terme désignant cette catégorie » (Charolles, 2002 : 185). À l'écrit le pronom de troisième personne n'est pas suffisamment autonome pour exister de lui-même. Une phrase telle que (59) ne peut être comprise qu'à partir du moment où l'interlocuteur raccroche le

[12] Terminologie de grammaire traditionnelle, cf. Riegel & *al.*, (1994)

pronom à une entité saillante, c'est-à-dire perçue dans la situation d'énonciation (à l'oral) ou fixée par le biais d'un nom propre ou d'une description à l'écrit :

(59) Elle a acheté une voiture.

En effet, comme le précise Charolles (2202 : 217) : « l'entité visée doit être en connexion […] avec un évènement [ou actant] saillant dans l'esprit des participants à l'échange » (voir aussi Kleiber, 1994). Ainsi, un pronom de troisième personne peut être anaphorique ou cataphorique puisqu'il nécessite la présence d'un référent pour être interprété (cf. chapitre 2).

D'autre part, les pronoms personnels sont liés à la prédication du verbe car ils réfèrent aux entités qui, dans la phrase, remplissent des fonctions grammaticales (sujet, complément d'objet) et par là même, assument les fonctions d'agent, d'objet ou de bénéficiaire. C'est le cas dans l'exemple (60) emprunté à Charolles (2002) :

(60) Il la lui a prêtée.

où les pronoms saturent par avance la structure actancielle appelée par le verbe (*ibid.*) : le pronom personnel *il* sujet réfère au rôle d'agent, *la* complément d'objet direct au rôle d'objet et *lui* complément d'objet second au rôle de bénéficiaire.

Les pronoms de dialogue

Les pronoms de dialogue (pronoms de 1e et 2e personne) sont ainsi nommés car ils désignent de manière exclusive les participants à une interaction verbale (discussion à l'oral ou dialogue à l'écrit). De ce fait, ils sont déictiques car interprétables uniquement dans la situation dans laquelle ils sont énoncés. Ils réfèrent aux locuteurs et interlocuteurs engagés dans l'échange, et uniquement à ceux-ci.

En effet, dans le cas du pronom de première personne *je*, on est tenté d'assimiler le *je* du dialogue (l'actant) à *la personne qui dit je* (une personne déterminée). Or, ce pronom n'est pas équivalent à la description définie *la personne qui prononce je*, comme l'explique Charolles (2002 : 198) :

> « *je* ne renvoie pas à une pure instance discursive, le pronom vise et atteint la personne en tant que telle, qui ne se réduit pas au fait qu'elle prononce actuellement *je*. Cette visée passe certes par le discours, mais le résultat attendu et obtenu n'est pas de désigner seulement le locuteur. *Je* en disant *je* manifeste qu'il est sujet de l'énonciation, mais il fait plus, il s'expose en tant qu'ego, en tant que moi ».

En outre, le *je* s'adresse à un interlocuteur désigné par le pronom de deuxième personne *tu*. Mais dans la bouche de ce *tu*, *tu* devient *je :* « Je n'emploie *je* qu'en m'adressant à quelqu'un qui sera dans mon allocution un *tu.* C'est cette condition de dialogue qui est constitutive de la personne car elle implique en réciprocité que *je* devient *tu* dans l'allocution de celui qui à son tour se désigne par *je* » (Benveniste, 1966 : 260). De ce fait, les pronoms de dialogue sont des expressions référentielles puisqu'ils renvoient à un référent de la situation dans laquelle ils sont produits mais ne peuvent être anaphoriques au sens où nous l'entendons dans le chapitre suivant. (cf. chapitre 2).

Synthèse des notions centrales du chapitre 1

Dans ce chapitre, nous avons tenté de définir et de distinguer les notions de cohérence, de cohésion et de continuité référentielle en les abordant sous différents angles.

Nous avons ainsi vu que la cohérence d'un discours (oral ou écrit), d'un point de vue psycholinguistique, implique des phénomènes cognitifs construisant une image mentale nécessaire à la compréhension de l'énoncé (Gernsbacher & Givon, 1995 ; Givon, 1995a ; Traxler & Gernsbacher, 1995). Du point de vue de l'encodeur, la production d'un discours nécessite donc plusieurs étapes pour façonner cette représentation mentale (code de surface, base de texte, modèle de situation, type de texte, niveau de communication), dont le décodeur doit reproduire le schéma en sens contraire pour accéder au sens. En outre, au niveau linguistique, ont été formulées des règles de « bonne formation » du texte. Il s'agit des 4 méta-règles de la cohérence (Charolles, 1978 ; Reinhart, 1980) : répétition, progression, non contradiction et relation. La cohérence peut donc s'appréhender aussi comme le passage d'un modèle au niveau cognitif à un message linéaire sous forme de discours au niveau linguistique (Fayol, 2002). Pour « mettre en forme » le discours au niveau linguistique, le locuteur/scripteur peut faire intervenir des marques linguistiques lui permettant de lier les énoncés entre eux, permettant ainsi la cohésion du texte : il s'agit de marques cohésives comme l'ellipse, la substitution, la conjonction et la cohésion lexicale (Halliday & Hasan, 1976).

La cohésion et la cohérence d'un texte sont aussi assurées par des éléments associés au statut de l'information véhiculée. La saillance par exemple, détermine la mise en exergue

d'une information par des procédés linguistiques d'ordre syntaxique (constructions syntaxiques, cf. Landragin, 2004), sémantique (type de caractère d'une entité, cf. Lyons, 1980 ; Pattabhiraman, 1993) ou pragmatique (progression thématique, cf. Combettes, 1983, Caron, 1989). La saillance constitue ainsi une aide à la résolution des anaphores. Cette notion nous servira à l'analyse de nos données. En outre, selon que l'information (ou un référent) est considérée comme nouvelle (non encore mentionnée dans le discours) ou ancienne (Clark & Haviland, 1977 ; Chafe, 1994), elle n'aura pas le même degré d'accessibilité (Levelt, 1989) pour le destinataire. Ceci engendre non seulement le fait qu'elle n'occupera pas la même place dans le discours (Lambrecht, 1994 ; Hickman, 1995), mais elle ne sera pas non plus réalisée par les mêmes formes linguistiques d'expressions référentielles (Prince, 1981 ; Ariel, 1988 ; Gundel & *al*., 2000).

Les expressions qui marquent la continuité référentielle sont nombreuses. Dans ce chapitre, nous nous sommes attachées à décrire de manière très succincte les principales formes typiquement rencontrées dans les corpus d'écrits scolaires, à savoir les SN et leurs déterminants (SN définis, démonstratifs, indéfinis), les pronoms, notamment personnels, ainsi que les noms propres. Tous ces éléments permettent au locuteur de construire la référence en exploitant les traits descriptifs des entités et la façon dont elles sont accessibles dans le contexte (Charolles, 2002). Ainsi, il est admis de considérer ces expressions linguistiques comme anaphoriques car on considère leur capacité à faire reconnaître un terme comme antécédent (Corblin, 1985).

Dès lors, nous avons pu constater que la notion d'anaphore est présente dans tous les domaines que nous venons d'aborder. Cette présence en filigrane paraît naturelle car, comme le montre Charolles (1995 : 3) :

> « [l'] anaphore assure des solidarités référentielles (coréférence, associativité, etc.) entre certains constituants des énoncés et […] donne naissance à des chaînes de référence. »

C'est pourquoi, dans le chapitre suivant, nous allons nous attacher à définir de façon plus précise ces deux concepts, essentiels à notre recherche, que sont l'anaphore et la chaîne de référence.

CHAPITRE 2 : Anaphores et Chaînes de Référence

L'anaphore est fondamentale pour assurer la cohésion, parce qu'elle permet d'assurer la continuité référentielle entre les mentions d'une même entité dans un discours et contribue ainsi à construire la référence.

De très nombreux travaux en linguistique ont été consacrés à l'anaphore. Deux grands types d'approches se dégagent : les approches qui privilégient l'angle textuel et celles qui envisagent l'anaphore comme un phénomène mémoriel. En 2.1., nous décrivons ces deux points de vue, ainsi que les différents types d'anaphores. À travers cette première partie du chapitre, nous montrons comment, loin d'être antagonistes, les deux approches sont complémentaires dans notre travail. En effet, dans la mesure où nous nous intéressons à la production écrite, nous avons besoin des deux approches car le rédacteur doit également prendre en compte les représentations du lecteur. En outre, ces approches permettent de décrire les phénomènes référentiels à la fois du point de vue du décodeur (réception) et de celui de l'encodeur (production).

Néanmoins, l'anaphore ne suffit pas à elle seule à exprimer toutes les mentions d'une entité dans un discours. La multiplication des mentions au même référent donne naissance à une chaîne, dont chacun des maillons est un marqueur référentiel de la même entité. Il s'agit d'une chaîne de référence. Ainsi, pour rendre compte de la continuité référentielle, nous avons besoin de la notion d'anaphore mais aussi de celle de chaîne de référence. Aussi, en 2.2., nous définissons cette notion à travers la présentation de ses éléments caractéristiques. Dans la mesure où notre corpus porte sur l'analyse de textes narratifs, nous présentons également la réalisation des chaînes de référence en fonction des différents genres textuels, et en particulier les travaux portant sur des textes narratifs.

2.1. L'anaphore

L'anaphore est une notion qu'il est difficile de définir car elle peut s'approcher sous divers angles. D'abord, l'anaphore peut s'étudier sous un angle textuel (ou purement linguistique). Dans ce cas, il s'agit principalement de distinguer un énoncé endophorique d'un

énoncé exophorique et donc, plus généralement, de distinguer l'anaphore de la coréférentialité et l'anaphore de la cataphore. La définition de ces notions fait l'objet de la section 2.1.1.

En 2.1.2., nous nous attachons à décrire différentes catégories d'anaphores qui nous sont utiles pour l'analyse de nos données. Cette revue de questions ne se veut pas exhaustive. Parmi les différents types d'anaphores que nous décrivons, nous nous attardons sur les anaphores nominales et ses sous-catégories : fidèles/infidèles mais aussi associatives et résomptives. Enfin, nous abordons les anaphores pronominales.

2.1.1. L'anaphore : une approche plurielle

Nous visons dans cette partie un bilan orienté vers deux approches de l'anaphore : l'approche textuelle et l'approche mémorielle.

2.1.1.1. L'anaphore comme phénomène textuel

La référence est une notion qui implique plusieurs éléments. Lorsqu'un interlocuteur fait d'une entité son objet de discours, il se limite rarement à la mentionner une seule fois. Il fait donc appel à des expressions qui reprennent l'entité dont il est question (ou *référent*) tout au long du propos et qui sont responsables de la continuité référentielle du texte (Apothéloz, 1995b, Charolles, 2002). Ces reprises du référent dans le discours s'appellent des *anaphores*.

L'anaphore est un rapport entre deux termes, plus précisément entre des expressions linguistiques permettant l'identification référentielle. Pour préciser son fonctionnement, reprenons un exemple et son explication tirés de Kleiber (1988 : 2).

(61) Alfred est saoul. *Il* a bu du schnaps.

Pour comprendre à qui renvoie le pronom personnel *il*, il faut trouver le référent de cette expression linguistique. Le pronom personnel, appelé *expression anaphorique* est identifié au moyen d'une autre expression linguistique (*Alfred*, le nom propre sujet de *est saoul*) présente dans le contexte linguistique antérieur. C'est pour cette raison qu'elle est nommée *antécédent.* Tesnière (1959) - et à sa suite Corblin, 1985 - préfèrent le terme de *source* à celui

d'*antécédent* car ce premier terme ne donne pas d'indication quant à la place qu'occupe la mention du référent dans le texte (avant ou après / à droite ou à gauche). En effet, *antécédent* implique que la mention précède l'anaphorique. Or, il est des cas où elle le suit, comme l'explique Tesnière (1959 : 86-87) :

> « Malheureusement, le terme *antécédent* a l'inconvénient de désigner le mot en question, non d'après sa nature, qu'on ne recherche même pas, mais d'après sa position, qui est sujette à toutes les variations que lui impose l'ordre linéaire de la chaîne parlée. »

Pour cette raison, nous préférons employer nous aussi le terme de *source.*

Revenons à notre exemple. *Alfred* est donc la source de l'expression anaphorique *il*. Comme cette expression anaphorique est déjà présente dans le contexte linguistique sous la forme de sa source, on peut dire qu'elle reprend une mention antérieure. Il y a de ce fait un phénomène de reprise. À ce titre, il y a coréférence entre l'expression référentielle et sa source. Il s'agit donc d'une anaphore coréférentielle. On peut illustrer la relation anaphorique par un exemple typique tel que celui de Corblin (1985) :

(62) Le lynx disparaîtra. *Il* est pourtant protégé.

Dans cet exemple, le pronom *il* ne peut s'interpréter que comme reprenant le syntagme nominal (SN) défini *le lynx* car ce segment du discours répond aux conditions d'interprétations qu'exige le pronom : les deux sont masculins et les deux sont singulier. *Il* est donc un pronom anaphorique dont la source est le SN défini *le lynx*.

Les situations où le référent d'un pronom anaphorique est présent dans le co-texte, comme celles que nous venons de voir, sont nombreuses. Mais il y a aussi des cas où le référent est localisé dans la situation d'énonciation (ou espace non discursif). C'est le cas par exemple lorsqu'un adulte s'adresse à un enfant qui s'approche trop près d'un chien :

(63) Attention, *il* est dangereux !

Dans cet exemple emprunté à Kleiber (1988 : 2), il n'y a pas de mention antérieure au pronom *il*. Ainsi, il faut différencier les énoncés (61), (62), dans lesquels la référence est dite *endophorique* ou textuelle, de l'énoncé (63) dans lequel la référence est appelée *exophorique* ou situationnelle. Pour plus de précisions concernant cette distinction, nous renvoyons à Maillard (1974), Halliday & Hasan (1976), Fraser et Joly (1980).

Lorsqu'il présente la référence endophorique, Kleiber (1988 : 3) définit l'anaphore comme :

> « (…) un processus de reprise où une expression anaphorique renvoie à un référent déjà mentionné dans le discours. Les expressions anaphoriques sont, par conséquent, en relation de coréférence avec leur antécédent ».

Néanmoins, la coréférentialité n'est pas une condition *sine qua non* de la réalisation de l'anaphore (Milner, 1982 ; Corblin, 1985 ; Kleiber, 1994). En effet, les expressions linguistiques coréférentielles qui ne tirent pas leur interprétation référentielle d'une expression antérieure en première position dans le discours, ne sont pas anaphoriques. En témoignent les exemples suivants annotés par l'auteur (Kleiber, 1988) :

(64) Je$_1$ suis venu te$_2$ voir et tu$_2$ m$_1$'as fermé la porte au nez.

(65) Alfred$_1$ était arrivé avec Céline$_2$. Céline$_2$ était gaie, Alfred$_1$ triste.

(66) Les vignes rapportent beaucoup aux viticulteurs$_1$, mais il faut dire que les viticulteurs$_1$ travaillent durement.

Ici, les pronoms personnels, noms propres et SN génériques en seconde position dans le discours (notés $_2$ à côté du mot concerné) sont en relation de coréférence avec les expressions en première position (notées $_1$). Mais ils ne tirent pas leur interprétation référentielle de ces expressions. Ils ne sont donc pas anaphoriques. En d'autres termes, ces expressions linguistiques sont liées car elles renvoient au même référent (coréférence) mais la première « mention » n'est pas nécessaire pour interpréter la seconde (non anaphorique).

À l'inverse, certaines expressions tirent leur interprétation référentielle d'une expression antérieure sans que celle-ci soit coréférentielle. C'est le cas notamment pour les anaphores associatives (67), les anaphores lexicales (68) ou l'anaphore « générique » (69) que l'on retrouve dans les trois exemples ci-dessous proposés par Kleiber (1988) :

(67) C'était un beau village. L'église était située sur une bute.

(68) Paul a tué trois lions, Alfred en a tué cinq.

(69) Un caniche m'a mordu. Ces chiens sont pourtant très gentils, en général.

Dans (67), l'église n'a pas été introduite une première fois de manière explicite et reprise ensuite par le syntagme nominal défini *l'église*. C'est au lecteur d'inférer que s'il y a

un village, il a une église. Il n'y a pas coréférence entre les expressions *un village* et *l'église.* Pourtant, le syntagme nominal *l'église* est bien une expression anaphorique.

De même pour (68), le référent *lion* n'est pas repris dans la deuxième partie de la phrase, comme ce serait le cas dans (68') :

(68') Paul a tué trois lions, Alfred a tué cinq lions.

C'est le pronom *en* qui fait office de reprise de la source.

Enfin dans (69), il n'y a pas non plus de coréférentialité car on passe du spécifique (*caniche*) au générique (*chien*). Les trois derniers exemples d'anaphores non coréférentielles montrent bien le fait que la source doit être prise en compte pour qu'il y ait interprétation référentielle de l'expression anaphorique, sans qu'il y ait obligatoirement relation de coréférence.

Nous l'avons vu, l'anaphore implique de mettre en relation la première et la deuxième expression pour identifier le référent de la seconde expression. Ainsi dans l'exemple (70) d'anaphore coréférentielle (emprunté à Reichler-Béguelin, 1988a), le référent est donné par l'environnement textuel (ou contexte linguistique) antérieur :

(70) J'ai rencontré *Pierre. Il* m'a parlé de toi.

Lorsque la source[13] apparaît en seconde position, comme en (71) (tiré de Reichler-Béguelin, 1988a) où le référent est livré par l'environnement textuel postérieur, on parle de *cataphore* (Reichler-Béguelin, 1988a ; Kleiber, 1988, 1991) :

(71) Quand je *l'*ai rencontré, *Pierre* a paru surpris de me voir

Même si la cataphore semble être le symétrique de l'anaphore dans le sens où la cataphore « anticipe » un élément linguistique quand l'anaphore le « reprend », il est cependant incorrect d'affirmer un tel parallélisme. En effet, la position de l'interprétant (avant/après ou à gauche de/à droite de) ne suffit pas pour désigner l'un ou l'autre de ces phénomènes de dépendance contextuelle[14].

[13] au sens terminologique de Tesnière (1959) et Corblin (1985). On voit ici tout l'intérêt d'employer le terme de *source* et non d'*antécédent.*

[14] Notre but n'étant pas ici de décrire en détail les spécificités de la cataphore, nous renvoyons en particulier à Kesik (1989) pour un panorama de la question.

Après ces clarifications, nous retiendrons comme définition de l'anaphore celle de Reichler-Béguelin (1988a : 17) :

> « On qualifie habituellement d'anaphorique un segment d'énoncé […] dont l'interprétation nécessite le recours à un autre segment, précédemment actualisé dans la chaîne discursive. […] L'anaphore représente un phénomène de dépendance interprétative entre deux unités, la seconde ne pouvant recevoir un sens référentiel sans avoir été mise en connexion avec la première […]. ».

Nous retenons cette définition parce que la notion de segment à laquelle recourt l'auteur peut à la fois signifier que le référent est donné directement par le contexte ou que le contexte donne les moyens d'identifier le référent. La notion de « dépendance interprétative » permet de faire la différence entre ce que Corblin (1995) appelle *l'identité d'interprétation* et *l'interprétation par reprise*. Dans le premier cas, il y a identité d'interprétation « si a et b reçoivent la même interprétation en vertu de règles qui ne doivent rien à leur proximité dans le même segment linguistique » (Corblin, 1995), c'est-à-dire par exemple un personnage prénommé *Paul* dénommé plus loin dans le discours par son lien de filiation *le fils de Marie*. Dans le deuxième cas, il y a interprétation par reprise « si un terme, b, exige pour être interprété, l'emprunt à un terme proche a d'un élément qui fixe l'interprétation de b » *(ibid.)*. Par exemple, le personnage prénommé *Paul* repris plus loin par le pronom personnel *il*.

La distinction est importante car elle montre bien que l'anaphore n'est pas le simple rappel d'un référent préalablement introduit dans le discours. En effet, il arrive souvent qu'il y ait des discordances morpho-syntaxiques, sémantiques et/ou référentielles qui interviennent entre l'anaphorique et sa source (Reichler-Béguelin, 1988a), comme le montrent les exemples déviants qui suivent (72, 73), empruntés à l'auteur :

> (72) Il trouva *trois champignons* et fut étonné car *ce champignon* était très rare. (Exemple de Corblin, 1985 : 386 repris par Reichler-Béguelin, 1988a)

Dans (72), l'emploi du SN démonstratif semble inapproprié au regard du référent qui est au pluriel *trois champignons*. Pourtant, on comprend que le passage au singulier signifie qu'on ne parle plus de ces trois champignons précisément, mais de la variété de champignon qui fait des champignons trouvés qu'ils sont rares.

Il peut arriver aussi que le référent de l'expression anaphorique ne figure aucunement dans le contexte discursif :

(73) Il fallait entre autres choses que cette pauvre enfant allât deux fois le jour puiser de l'eau à une grande demi-lieue du logis, et qu'elle en rapportât plein une grande cruche. Un jour qu'elle était à *cette fontaine*, il vint à elle une pauvre femme qui la pria de lui donner à boire. (Perrault, *Contes*, Folio, 165, exemple de Reichler-Béguelin, 1988a)

Ici, *cette fontaine* ne réfère pas directement à une occurrence qui aurait déjà été mentionnée, mais fait appel au savoir du locuteur qui infère le fait que si la jeune femme va puiser de l'eau, c'est qu'il y a une fontaine ou un puits quelque part. Cet exemple déviant peut être compris à condition que le lecteur fasse un calcul interprétatif.

On peut rencontrer enfin des cas où le référent est un être sujet à des modifications et dont la description évolue (Reboul, 1989 ; Kleiber, 1991). C'est le cas par exemple du poulet rôti de Brown et Yule (1983) :

(74) Kill and active, plump chicken. Prepare *it* for the oven, cut *it* into four pieces and roast *it* with thyme for 1 hour.

(74') Tuez un poulet vif et bien gras. Préparez-*le* pour le four, coupez-*le* en quatre et rôtissez-*le* [avec du thym] pendant 1 heure. (traduction de Reboul, 1989)

Dans cet exemple, le poulet de départ, *vif et bien gras*, n'est plus tout à fait le même à l'arrivée puisqu'il est *coupé en quatre* et *rôti*. Il est bien question du même être (il a la même identité), mais sa description a évolué, comme le signalent Brown & Yule (1983 ; traduit par Reboul, 1989) :

« un lecteur qui remonterait simplement la chaîne endophorique et qui substituerait l'expression *un poulet vif et bien gras* à *le* dans le dernier syntagme, aurait échoué à comprendre le texte ».

De cette manière, l'attribution d'un référent à une expression anaphorique passe par la représentation mentale que s'en fait l'interlocuteur et non plus seulement par le discours préalable (Reboul, 1989).

Les derniers exemples nous conduisent à approcher l'anaphore comme un phénomène mémoriel.

2.1.1.2. L'anaphore comme phénomène mémoriel

L'anaphorisation d'une entité est une opération qui insère un objet dans un domaine pour l'extraire en vue d'une focalisation particulière (Fall, Vignaux & Boivin, 1994), en faisant appel au co-texte, comme on le voit dans l'exemple (75), et/ou en faisant appel aux savoirs partagés par les interlocuteurs, comme le montre l'exemple (76). Les deux exemples sont empruntés aux auteurs.

(75) J'ai acheté *un gâteau* en ville. *Ce gâteau* m'a coûté une fortune.

(76) La maison de la rue Fraser… Cette ancienne école rénovée…

En (75), le SN démonstratif *ce gâteau* de la seconde phrase fait référence au SN indéfini *un gâteau* présent dans la phrase précédente. Dans (76), le SN démonstratif *cette ancienne école rénovée* fait appel au savoir partagé des interlocuteurs : pour comprendre la référence, il faut savoir que *la maison de la rue Fraser* est une ancienne école.

L'approche mémorielle met en avant le critère de la saillance préalable pour définir l'anaphore, comme l'indique Kleiber (1994 : 25) :

> « L'anaphore devient un processus qui indique une référence à un référent déjà connu par l'interlocuteur, c'est-à-dire un référent « présent » ou déjà manifeste dans la mémoire immédiate. »

Les exemples suivants (77) et (78), empruntés à Kleiber (1994), illustrent cette notion de saillance préalable :

(77) *Paul* est sorti. *Il* avait trop chaud.

(78) Attention ! Ne t'approche pas. *Il* est dangereux.

Le pronom *il* est anaphorique puisque le référent est déjà saillant : dans (77), par la mention antérieure du nom propre *Paul*, dans (78), par la situation d'énonciation elle-même.

L'anaphore peut aussi être considérée comme un phénomène de rappel informationnel où sont susceptibles d'intervenir non seulement « le savoir construit linguistiquement par le texte lui-même », mais aussi « les contenus inférentiels qu'il est possible de calculer à partir des contenus linguistiques » (Reichler-Béguelin, 1988a : 18). Sur ce dernier point en effet, il arrive souvent que l'identification d'un référent (ou sa construction) soit à la charge des

compétences inférentielles du destinataire, comme dans les exemples suivants (tirés de Reichler-Béguelin, 1989, que nous commentons en page suivante) :

(79) Atterrir sur la place Rouge ! Certainement *le jeune pilote* est fou… (Courrier de la *Suisse*, début de texte)

(80) Au crépuscule d'une belle journée de printemps, ce cultivateur de la région de Madarounfa, au sud du Niger, examine une dernière fois sa future récolte avant de rentrer au village. Vendu un bon prix, *ce coton* devrait lui permettre d'acheter le mil qui manquera, cette année encore, pour la soudure.

Dans la phase d'encodage d'un discours, tout référent potentiel passe par une phase de mise en mémoire, appelée selon les auteurs *univers de discours* (Lyons, 1980), *mémoire discursive* (Berrendonner, 1986 ; Reichler-Béguelin, 1988a ; 1989), *modèle contextuel* (Bosch, 1983), *modèle de discours* (Cornish, 1990) ou encore *focus* (Garrod & Sanford, 1982). Comme l'explique Johnsen (2014) au sujet de la mémoire discursive, le discours a pour fonction l'élaboration d'un ensemble évolutif de représentations partagées par les interlocuteurs. Le référent, présent dans cette mémoire discursive, est ensuite réextrait anaphoriquement dans un autre segment du discours, avec, selon les cas, plus ou moins de modifications (cf. les différents types d'anaphores). Or, lorsque l'encodeur produit son discours, il a en tête des contenus mnésiques d'origine langagière ou perceptuelle évidents pour lui, des éléments linguistiques qui renvoient à ses représentations psychiques dont la référence n'a pas nécessairement besoin d'être actualisée dans le discours de son point de vue. Ces représentations de l'encodeur recouvrent la « situation ambiante » (Wilmet, 1986 : 165) de production du discours[15]. Il s'agit alors de « déictique du locuteur » (Reichler-Béguelin, 1989 : 310) car ces représentations sont toujours explicites pour l'encodeur étant donné qu'elles viennent de lui-même. Mais elles sont moins évidentes pour le décodeur/destinataire qui, lui, n'a pas accès à toutes les conditions de production du discours. Celui-ci doit donc accéder à l'information anaphorisée par ses propres compétences inférentielles, en la construisant ou reconstruisant par conjectures et en s'appuyant pour cela sur le co-texte disponible. Ce calcul inférentiel s'appuie donc sur les éléments déjà saillants du texte, de la situation immédiate et du savoir d'arrière-plan présumé partagé, éléments qui permettent d'accéder au référent visé (Reichler-Béguelin, 1988a, 1989 ; Kleiber, 1994). C'est le cas dans l'exemple (79) vu plus haut, où c'est le verbe *atterrir* qui doit faire inférer au décodeur qu'il y a certainement un avion (si ce n'est un engin volant) et à ce titre, un pilote qui le conduit. De la même manière dans (80), le décodeur doit inférer grâce à son savoir que la récolte dont il

[15] Nous renvoyons à la théorie de l'énonciation pour plus de détails (Wilmet, 1986)

est question, puisqu'elle a lieu au Niger, a de forte chance d'être composée de coton. Précisons que le calcul interprétatif s'effectue de façon différente selon la posture du lecteur. Ce dernier aura tendance à être plus coopératif lorsqu'il lit, pour son intérêt personnel, un écrit de presse ou un texte littéraire par exemple, plutôt que lorsqu'il est dans une posture de correcteur. Dans ce cas, il aura beaucoup plus tendance à vouloir faire rentrer dans la norme un texte d'apprenant par exemple (Reichler-Béguelin, 1988a, 1989).

L'approche cognitiviste/mémorielle met en avant le statut cognitif des référents (Fossard, 2008). Elle est critiquée par certains linguistes (Cornish, 1999, 2001 ; De Mulder, 2000 ; Kleiber, 1994, 2001) car elle ne tiendrait compte que de la représentation mentale pour accéder au référent, la prédication anaphorique n'étant pas considérée, ni son éventuel aspect prédictif concernant l'orientation discursive à venir. De ce fait, son aspect purement « identificatoire » est reproché car c'est seulement le contexte antérieur qui est pris en compte dans l'analyse. Toutefois, la perspective mémorielle est intéressante lorsque l'on se place dans une perspective d'analyse en réception *vs* en production. En effet, le point de vue traditionnel fait de l'anaphore un phénomène textuel de reprise de l'information et se place uniquement du point du vue du destinataire-décodeur, autrement dit, du point de vue de la réception. Ainsi, pour être compris par son interlocuteur (ou décodeur), le locuteur (ou encodeur) doit fournir le maximum d'informations pertinentes dès l'apparition de chaque référent potentiel, sous forme linguistique. Les approximations, ajustements et recours simultanés à de l'information extra-linguistique sont *a priori* à éviter au maximum (Reichler-Béguelin, 1989).

Nous venons de le voir, l'anaphore peut aussi être envisagée comme un recours au savoir partagé (ou *mémoire discursive*), phénomène qui fait justement appel à de l'information extra-linguistique. Si l'on considère l'anaphore aussi comme un phénomène mémoriel de rappel de l'information, « censée incluse dans le savoir partagé par les partenaires de l'interlocution » (Reichler-Béguelin, 1988a), elle permet de décrire aussi bien la procédure de décodage (réception) que celle de l'encodage (production) et c'est ce qui nous intéresse particulièrement dans notre travail.

L'approche mémorielle de l'anaphore montre également que la conception textuelle n'est pas tout à fait correcte dans l'image qu'elle donne de l'anaphore (Kleiber, 1991). En effet, selon celle-ci, l'interlocuteur part de l'expression anaphorique pour remonter « en arrière » dans le texte jusqu'à localiser sa source Or, l'approche mémorielle montre que l'identification de l'expression anaphorique se fait par la représentation mentale construite à

partir du discours (Reboul, 1989) : c'est au moment où l'interlocuteur rencontre l'expression anaphorique qu'il va chercher dans sa mémoire discursive la bonne source et non pas revenir en arrière dans le texte (ou le discours) pour y trouver l'endroit exact où est localisée la source. Toutefois, ce retour en arrière s'effectue tout de même lorsque la résolution de l'expression anaphorique est problématique, dans le cas de plusieurs référents concurrents par exemple.

2.1.1.3. L'anaphore : un phénomène complexe

Nous venons de le voir, dans l'analyse de l'anaphore, deux grandes tendances se sont mesurées. Il ne s'agit pas ici de décider si l'anaphore doit être considérée comme un phénomène textuel dans lequel le texte est le seul indicateur de la référence. Il est admis aujourd'hui que cette approche textuelle n'est pas la plus pertinente (cf. Kleiber, 1994). Il ne s'agit pas non plus de déterminer si l'anaphore doit être approchée exclusivement comme un phénomène mémoriel où l'environnement extra-linguistique donne accès à la référence. Dans nos travaux, nous prenons le parti de voir l'anaphore comme un phénomène complexe, qui peut s'approcher sous divers angles, comme le dit Kleiber (1991 : 11) :

> « environnement extra-linguistique immédiat comme texte sont […] deux sources d'alimentation possibles de la *mémoire discursive* (Berrendonner, 1986 ; Reichler-Beguelin, 1988 a et b, 1989), qui donnent lieu, après introduction du référent, à un processus de référence anaphorique. »

Les deux approches de l'anaphore sont complémentaires, car selon Reboul (1989 : 88) « dès que l'on veut récupérer l'interprétation complète d'un énoncé, le recours à une analyse strictement linguistique ne saurait suffire ». Comme l'indique Kleiber (1994 : 13) :

> « les principes pragmatiques cognitifs n'interviennent pas directement [dans le processus d'interprétation référentielle], mais s'articulent en aval aux règles sémantico-référentielles propres à chaque type d'expression. »

Ainsi, quel que soit le texte envisagé, l'identification référentielle de la « bonne » source passe à la fois par le texte lui-même et par le savoir extra-linguistique.

2.1.2. Les différentes catégories d'anaphore

Il existe plusieurs catégories d'anaphores, et plusieurs façons de les classer. Notre but ici n'est pas de faire une revue de ces catégorisations, ni même d'être exhaustive dans la présentation de tous les types d'anaphores existants, mais de nous attacher à décrire ceux qui sont rencontrés généralement dans les corpus de textes d'élèves. De ce fait, nous prenons le parti traditionnel de différencier les anaphores en deux grandes catégories : anaphores nominales et anaphores pronominales, comprenant chacune leurs sous-catégories spécifiques.

2.1.2.1 Les anaphores nominales

Anaphores fidèles et infidèles

L'anaphore nominale (ou anaphore par nominalisation) « transforme le procès dénoté par une proposition antérieure en référent, c'est-à-dire en un objet individué » (Apothéloz, 1995a : 37). En d'autres termes, il y a anaphore nominale lorsqu'un objet cité dans une proposition fait à nouveau l'objet d'un discours dans une proposition ultérieure, qu'il s'agisse de rappeler le contenu propositionnel, comme dans (81) avec *la vente*, ou bien l'acte langagier accompli au moyen de l'énonciation de ce contenu, tel que *cette interrogation* dans (82) (Apothéloz, 1995a, exemples de l'auteur) :

(81) Les archives du cinéaste Abel Gance […] seront vendues aux enchères à Drouot les 3 et 4 mars. *La vente* comprendra une centaine de scénarios manuscrits.

(82) Convient-il de situer les fonctions dans l'ordre figuratif et les fonctionnements dans l'ordre figural ? Ou l'inverse ? *Cette interrogation* est appelée à durer.

Les anaphores fidèles et infidèles font partie des anaphores nominales. Elles sont appelées ainsi selon les relations sémantiques qu'entretiennent les unités lexicales impliquées (Reichler-Béguelin, 1988a, 1989 ; Apothéloz, 1995a). Leur but est de rappeler de l'information déjà verbalisée, c'est-à-dire « linguistiquement » fournie dans l'énoncé.

L'anaphore fidèle (ou stricte selon les terminologies) est appelée ainsi car elle reprend le nom tête de la source mais non le déterminant le précédant. Par exemple dans (83), le nom tête *vache* est repris mais le déterminant change d'une phrase à l'autre :

(83) Une vache… Cette/la vache… (exemple de Reichler-Beguelin, 1989)

L'anaphore infidèle, quant à elle, emploie un nom pour la forme introductive et un autre pour la forme de rappel. Les exemples qui suivent, empruntés à Reichler-Beguelin, (1988a), montrent que l'anaphore infidèle reprend la source par divers procédés.

L'anaphore hyperonymique est basée sur une relation hyponyme/hyperonyme, comme le montre (84) :

(84) Un soldat… cet/l'homme

Le premier terme employé pour désigner le référent est un hyponyme (*un soldat*). Il est repris ensuite par un hyperonyme ou un nom générique (*l'homme*).

L'anaphore infidèle peut aussi se traduire par une anaphore dans laquelle se retrouvent un nom et son synonyme ou quasi synonyme. C'est le cas dans (85), où la source est désignée par *un soldat* et l'anaphorique par un synonyme, *le militaire* :

(85) Un soldat… ce/le militaire

Enfin, dans la catégorie des anaphores infidèles, il y a également les anaphores conceptuelles. Il s'agit d'anaphores qui « rendent compte de la cohérence d'un raisonnement, synthétisé et rappelé, pour renforcer la force démonstrative du propos » (Marin, 2007 : 33), comme l'illustre l'exemple (86) :

(86) Il s'est décidé pour la médecine. Ce/le choix est judicieux.

L'anaphorique *ce/le choix* renvoie à la proposition précédente dans son ensemble. Cette proposition fait office « d'objet de pensée » (Moirand, 1975 : 7) qui est ensuite nominalisé sous la forme d'un SN, ici par *le/ce choix*.

Dans tous les cas, que ce soient les anaphores fidèles ou les anaphores infidèles, les anaphores nominales se caractérisent par leur « homogénéité catégorielle » (Milner, 1982 : 32) car la première mention du référent, comme l'anaphorique, se présente sous la forme de syntagmes nominaux (SN). Le seul élément qui indique la différence est l'emploi du défini qui marque la reprise anaphorique ou bien celui de l'indéfini qui introduit le référent.

Anaphore associative

L'anaphore associative est une bonne illustration d'un certain type d'anaphore qui ne « reprend » aucun élément explicitement mentionné ailleurs dans l'énoncé ou le contexte. L'anaphore ici, ne s'accompagne pas de coréférence mais se caractérise par le fait qu'elle « rappelle » une information accessible, présentée comme connue, mais non directement fournie par le contexte linguistique et sans indication de rapport aux autres référents ou aux autres informations explicitement formulées (Blanche-Benveniste & Chervel, 1966 ; Reichler-Béguelin, 1989 ; Charolles, 1990 ; Kleiber, 1994 ; Apothéloz, 1995a).

Pour donner un sens référentiel aux éléments de ce type d'anaphore, il faut recourir au savoir partagé (Berrendonner, 1983, 1986 ; Reichler-Béguelin, 1989) d'une communauté donnée qui autorise les raisonnements implicites. Ce savoir partagé est « relatif aux rapports de type méréologique et métonymique qui existent entre les réalités dénotées, dont certaines peuvent ainsi devenir objets de discours sans avoir été introduites au préalable » (Reichler-Béguelin, 1988a : 19, l'exemple est tiré de l'auteur) :

(87) Paul vient d'acheter *une maison*. *Le toit* et *les murs* sont en bon état, mais *les fenêtres* et *les volets* sont à refaire. *L'ancien propriétaire* lui a indiqué un bon menuisier.

Dans (87), les éléments en italiques sont des expressions anaphoriques non coréférentielles du SN *maison.* En effet, s'il y a une maison, elle est certainement composée de murs, d'un toit, de fenêtres et de volets (quoique ce dernier élément dépende du pays dans lequel la maison est construite). L'exemple (87) n'est pas sans rappeler celui de Kleiber (1988, cf. 2.1.1.1.) repris ci-dessous, où l'anaphore associative est matérialisée par *l'église* appartenant au *village* :

(67) C'était un beau village. L'église était située sur une bute.

Anaphore résomptive

L'anaphore résomptive est un cas particulier. En effet, contrairement à la plupart des anaphores, l'anaphore résomptive se manifeste par une expression démonstrative à travers laquelle le référent visé n'est pas un élément linguistique mais une représentation mentale. Son objectif n'est pas de référer à une relation de coréférence, une entité « objective » du discours (un personnage par exemple) mais à un contenu posé ou présupposé et d'instituer un cadre interprétatif (Kara & Wiederspiel, 2011). Pour illustrer l'anaphore résomptive, nous reprenons un exemple de Himmelman (1996) cité par Guillot (2006) :

(88) … puis il part… et c'est la fin de *cette histoire*, mais après… il revient vers le paysan…

Dans (88), le SN démonstratif *cette histoire* est anaphorique car les informations contenues dans le contexte précédent sont nécessaires à son interprétation. Toutefois, ce type d'anaphore réfère au contenu discursif d'un ensemble de propositions venant d'être énoncées, de quelques propositions à un discours entier (Maillard, 1974), sans que ce contenu n'ait été désigné auparavant par une expression référentielle particulière (Guillot, 2006). Ce sont les propositions énoncées auparavant qui font de ce discours une histoire.

2.1.2.2. Les anaphores pronominales

De manière générale, une anaphore pronominale se caractérise le plus souvent par la présence d'un référent sous la forme d'un nom (SN défini ou indéfini) et d'un anaphorique sous la forme d'un pronom de troisième personne. Pour illustrer cela, nous empruntons un exemple à E.T.A. Hoffman (89) :

(89) *La jeune et ravissante Giacinta Soardi* repoussa la somptueuse robe d'épais satin rouge dont *elle* cousait diligemment la garniture, et, du haut de sa fenêtre, regarda d'un air contrarié l'étroite ruelle désespérément vide. (E.T.A. Hoffman, *Princesse Bambilla,* notre exemple)

Le référent est la description définie *la jeune et ravissante Giacinta Soardi*, anaphorisée sous la forme du pronom personnel *elle*.

Dans l'anaphore pronominale : « la paire anaphorique est […] ouvertement hétérogène, du point de vue catégoriel » (Milner, 1982 : 32). C'est le nom anaphorisé qui apporte la référence au pronom. L'anaphore pronominale s'applique de fait à un segment de l'énoncé (Reichler-Béguelin, 1988a ; Reboul, 1989).

Anaphore sur syllepse

Au sens traditionnel, la syllepse est une figure de rhétorique qui consiste à employer un mot dans deux sens différents. Lorsqu'elle apparaît dans le contexte d'une anaphore pronominale, elle engendre des modifications du genre et du nombre grammatical de l'expression référentielle, comme dans les exemples (90) et (91) empruntés à Apothéloz (1995a) :

(90) Une femme infidèle, si elle est connue pour telle de la personne intéressée, n'est qu'infidèle ; s'*il* la croit fidèle, elle est perfide (La Bruyère, *Des femmes*).

(91) A la cantine, *elles* se sont plaintes qu'il fallait gratter le four.

Dans (90), *la personne intéressée,* expression référentielle de genre féminin, désigne le référent de la proposition repris par le pronom masculin *il*, suggérant ainsi que le référent du syntagme nominal défini *la personne intéressée* est de sexe masculin. L'accord en genre et en nombre se fait donc sur des bases référentielles et non grammaticales. De même dans (91), où le pronom personnel féminin pluriel *elles* peut être interprété comme référant à un syntagme qui pourrait être *les femmes de la cantine*. Dans ce cas, on passe d'une référence à une entité collective à une référence à la classe des individus que recouvre cette entité (Apothéloz, 1995a). Toutefois, Apothéloz (*ibid.*) précise que ce dernier exemple pourrait aussi être interprété comme une anaphore associative, les femmes de la cuisine pouvant être considérées comme « appartenant » au lieu dénommé *cuisine*. Le contexte a donc toute son importance dans l'interprétation.

Au terme de cette première partie consacrée à l'anaphore, nous pouvons décrire cette notion comme un rapport entre deux expressions linguistiques permettant l'identification référentielle. Selon l'approche textuelle, l'anaphore est le rappel d'un référent préalablement établi dans le discours. L'approche mémorielle, quant à elle, voit l'anaphore comme le rappel d'un référent déjà connu par l'interlocuteur car identifiable grâce aux savoirs partagés. Ces deux approches sont deux moyens complémentaires d'alimenter la mémoire discursive et d'accéder ainsi à l'identification référentielle (Kleiber, 1991). La référence anaphorique se traduit par différents types d'anaphores, que l'on peut classer en deux grandes catégories : les anaphores nominales et les anaphores pronominales.

2.2. Les chaînes de référence

Ainsi que nous venons de le voir, les expressions référentielles forment des paires, appelées *anaphores*. L'une se situe dans une première proposition dans laquelle elle occupe le rôle de source et l'autre se situe dans une proposition ultérieure.

Or, il n'est pas rare, dans un texte, de trouver de nombreuses expressions coréférentielles qui ont toutes pour source la même entité. Chaque expression référentielle est considérée comme un maillon d'une chaîne appelée *chaîne de référence*. La notion de chaîne est définie par Corblin (1995 : 123), à partir des travaux fondateurs de Chastain (1975), comme : « la suite des expressions d'un texte entre lesquelles l'interprétation construit une relation d'identité référentielle ».

Comme l'indique Schnedecker (1997), les *chaînes de référence* impliquent donc une complexité plus grande que la seule anaphore car elles nécessitent de manier des expressions référentielles en plus grand nombre.

2.2.1. Les chaîne de référence : définition

On ne peut définir précisément ce qu'est une chaîne de référence sans s'attacher à décrire les différentes caractéristiques de ses maillons, parmi lesquelles leur nombre, la distance entre eux, leur nature. La notion de chaîne n'est valable qu'à partir du moment où le nombre des expressions référentielles sur lesquelles on travaille est supérieur à deux unités, comme l'indique Corblin (1987a : 7) : « *Chaîne* permet de dépasser les contextes de simple succession de deux termes auxquels se limite le plus souvent le linguiste qui sort du domaine phrastique. » Cette précision apportée par Corblin est nécessaire car elle permet de différencier la notion d'anaphore de celle de chaîne de référence. Pour cette raison, à l'instar de Schnedecker (1997), nous parlerons de « chaîne » de référence à partir du moment où on compte au moins *trois expressions coréférentielles.* À l'inverse de l'analyse des anaphores qui peut s'effectuer à partir d'exemples construits, l'analyse des chaînes de référence s'appuie principalement sur des exemples attestés. Pour rendre compte de la notion de *chaîne de* référence, nous nous appuyons sur Schnedecker (1997).

Le nombre de maillons, s'il distingue l'anaphore de la chaîne, ne suffit pas à lui seul à spécifier une chaîne de référence car les chaînes peuvent présenter des caractéristiques diverses, comme le montrent les exemples suivants, empruntés à Schnedecker (1997) :

(92) Le lendemain, *Bony* resta au Trou du Pied-Bot. *Il* fit la lessive et prit la peine de cuisiner quelque chose qui ressemblait à du ragoût. *Il* remonta le sentier jusqu'au sommet de la falaise, le fusil sous le bras, espérant voir un kangourou car *il* avait remarqué des empreintes récentes dans le trou. *Il* passa

deux heures à examiner les anciens journaux et documents qu'*il* avait pris dans la vieille valise, sous le lit de Lonergan. (A. Upfield, L'homme des tribus).

(93) En homme qui ne craint ni le froid ni la neige, *Michel Strogoff* eût préféré voyager par la rude saison d'hiver, qui permet d'organiser le traînage sur toute l'étendue du parcours. Alors les difficultés inhérentes aux divers genres de locomotion sont en partie diminuées sur ces immenses steppes nivelées par la neige. Plus un cours d'eau à franchir. Partout la nappe glacée sur laquelle le traîneau glisse facilement et rapidement. Peut-être certains phénomènes naturels sont-ils à redouter, à cette époque, tels que permanence et intensité des brouillards, froids excessifs, chasse-neige longs et redoutables, dont les tourbillons enveloppent quelque fois et font périr des caravanes entières. Il arrive bien aussi que des loups, poussés par la faim, couvrent la plaine par milliers. Mais mieux eût valu courir ces risques, car, avec ce dur hiver, les envahisseurs tartares se fussent de préférence cantonnés dans les villes, leurs maraudeurs n'auraient pas couru la steppe, tout mouvement de troupes eût été impraticable, et *Michel Strogoff* eût plus facilement passé. Mais *il* n'avait à choisir ni *son* temps ni *son* heure. Quelles que fussent les circonstances, *il* devait les accepter et partir.

Telle était donc la situation que *Michel Strogoff* envisagea nettement, et *il* se prépara à lui faire face. (J.Verne, *Michel Strogoff*)

Bien que le nombre de maillons (signalés en italique) soit largement supérieur à deux dans les deux extraits précédents, l'effet produit en termes de « solidité » de la chaîne de référence (ou *texture* selon Halliday & Hasan, 1976 ; cf. chapitre 1, section 1.3.) n'est pas le même. En effet, dans (93), la deuxième mention du référent *Michel Strogoff* n'apparaît qu'une dizaine de lignes après la première. Alors que dans (92), toutes les mentions du référent sont très proches les unes des autres, quasiment une dans chaque phrase. La distance qui sépare chaque maillon de la chaîne (ou *distance inter-maillonnaire*) a une incidence sur la substance ou la consistance de la chaîne toute entière. Dès lors, nous pouvons nous interroger, à l'instar de Combettes & Tomassone (1988 : 14), sur la manière de mesurer cette distance :

« Délicat nous semble être le problème de la distance qui sépare la première occurrence du référent et le groupe « donné » ; à partir de quel moment va-t-on considérer qu'un élément, à l'écrit comme à l'oral aura disparu de la conscience du lecteur ? Faut-il estimer qu'une fois introduite dans le texte une unité est définitivement « donnée », même si elle n'est plus mentionnée pendant plusieurs pages, voire pendant plusieurs chapitres ? (…) Peut-on mesurer cette « distance », ce temps pendant lequel le référent demeure présent à l'esprit, en termes quantitatifs, s'agit-il de temps de lecture, de nombre de phrases ? »

En psycholinguistique, la distance inter-maillonnaire est évaluée en termes de temps de lecture. En ont été déduits deux facteurs de fonctionnement de la coréférenciation : le lecteur a tendance à prendre comme coréférent le syntagme le plus proche (*stratégie de*

distance minimale) et le temps de lecture augmente quand l'anaphorique est séparé de sa source par deux phrases[16].

En linguistique, il est difficile de donner une unité de mesure stricte sur laquelle s'appuyer pour déterminer à partir de quelle distance entre deux maillons une chaîne est rompue ou non. Que ce soit en termes typographiques (nombre de caractères ou de lignes entre deux mentions), morphologiques (nombre de syllabes ou de mots) ou configurationnels (nombre de phrases ou de paragraphes), les unités constituant l'intervalle séparant deux expressions coréférentielles restent « dans les faits trop malléables pour être véritablement opérantes » (Schnedecker, 1997 : 39). La délimitation d'une chaîne reste donc floue. À ce sujet, Schnedecker (1997) emploie le terme de *bornage*, pour désigner les limites d'une chaîne, composées d'un côté d'une expression linguistique qui en marque le début et de l'autre d'une expression linguistique qui en marque la fin. Toutefois, les chaînes de référence interagissent avec les autres plans de l'organisation textuelle : le paragraphe semble avoir une influence sur le bornage des chaînes de référence car il produit des effets sur l'interprétation des expressions référentielles. En cela, il fait ressortir deux paramètres qui conditionnent la cohésion référentielle inter-paragraphique (*ibid.* : 41). Le premier est celui de la densité des expressions coréférentielles à l'intérieur d'un même paragraphe : plus elles sont nombreuses, plus elles « renforcent » la chaîne et se retrouvent plus facilement d'un paragraphe à l'autre. Le second paramètre est celui de la répartition des expressions coréférentielles. En effet, meilleure est leur répartition dans le paragraphe, meilleures sont les transitions référentielles. Aussi, comme le résume Schnedecker (1997 : 44) :

> « l'existence de ces interactions [*ndlr* : entre le paragraphe et les maillons d'une chaîne] signifie-t-elle peut-être, tout simplement, que les questions de la distance et de l'unité ne doivent pas se poser en soi […] mais qu'elles doivent être corrélées avec l'usage qui est fait des expressions référentielles. »

S'il est difficile de délimiter le bornage d'une chaîne de référence en mesurant la distance qui sépare ses maillons, nous pouvons en revanche nous appuyer sur la nature des expressions référentielles. En effet, suivant Schnedecker (1997), la redénomination d'un référent par un nom propre indique un changement de point de vue et induit une rupture thématique car elle « signalerait qu'il va être dit, à propos du référent, des contenus de nature distincte » (*ibid.* : 186). De ce fait, le nom propre débute une nouvelle chaîne.

[16] Pour une revue de ces travaux, nous renvoyons à Schnedecker (1997).

D'autre part, la *texture* d'une chaîne de référence tient aussi à la nature des maillons qui la composent. Aussi, on peut se demander si toutes les expressions référentielles peuvent être considérées en tant que maillon d'une chaîne de référence. Pour Charolles (1988b : 8), seules les formes linguistiques devraient être prises en compte dans le comptage des maillons :

> « Seules peuvent appartenir (donner lieu à) une chaîne des expressions employées référentiellement, c'est-à-dire toutes et rien que les expressions nominales (ou pronominales) permettant d'identifier un individu (un objet de discours) quelle que soit sa forme d'existence (personne humaine, évènement, entité abstraite). »

Cette définition semble stricte car elle ne tient pas compte des éléments textuels en mesure d'informer le destinataire sur le référent, tels que les appositions, les anaphores zéro ou les phénomènes d'accords visibles comme dans (94), exemple de Schnedecker et Landragin (2014) souligné par eux :

> (94) Drew Barrymore « Les amours à distance c'est l'histoire de ma vie. »
>
> Né<u>e</u> dans les paillettes du 7e art
>
> **Elle** *boit* **son** thé glacé à la paille et Ø *répond* du tac au tac, *en ponctuant* régulièrement **ses** phrases du fameux « F Word ». **Drew Barrymore**, né*e* dans les paillettes du 7e art, *a* le vocabulaire et l'énergie d'une New-Yorkaise. Plus menu*e* que prévu – en slim gris, T-shirt Mickey Mouse et derbys noirs -, le teint naturel et le sourcil sérieux, **elle** *est* aux antipodes de l'image que l'on se faisait d'**elle**. Une hippie-chic frivole et bouillonnante ? Pas seulement. **Drew**, *l'aventurière extravertie capable de parcourir le globe avec sa meilleure amie (Cameron Diaz) pour une émission écolo, ou de montrer ses seins à David Letterman en plein talkshow télé*, **s**'est visiblement assagi*e* […] » (*Elle,* 20.08.10).

Dans cet extrait d'article de magazine, les expressions linguistiques en gras sont les maillons les plus communément admis. Il s'agit d'expressions nominales (SN, nom propre) et de pronoms permettant d'identifier l'individu dont il est question (Drew Barrymore). Les expressions en italiques sont pourtant autant d'informations sur le référent : appositions (dernière phrase en italique par exemple), phénomènes d'accords (*née*, *menue*, …), anaphore zéro (Ø).

Ainsi, on parle de *chaîne de référence* lorsqu'il y a au moins trois expressions coréférentielles ou *maillons*. Si le nombre de maillons suffit à déterminer une chaîne de référence, leur nature et la distance qui les sépare a une incidence sur l'effet produit sur le lecteur ou *texture*.

2.2.2. Les relations entre chaînes de référence

Les chaînes de référence sont donc constituées de plusieurs maillons et se déroulent tout au long d'un texte. Or, Schnedecker (1997) fait remarquer que ce déroulement semble difficilement prévisible, d'autant plus lorsque plusieurs chaînes référant à des entités différentes se déroulent en même temps. L'auteur relève différents types de relations entre les chaînes, que nous allons illustrer par des exemples qui lui sont empruntés (Schnedecker, 1997, exemples 95 à 100).

Tout d'abord, il peut s'agir de relations plutôt « simples » comme les relations de *succession* qui voient une chaîne s'interrompre quand en débute une autre :

> (95) Johannesburg, années 70. **Un professeur d'Histoire** vit une vie de famille sans histoire, entouré d'affection des **siens** et aveugle aux problèmes politiques et sociaux engendrés par l' « apartheid », régime ségrégationniste qui sévit en Union sud-africaine. **Cet homme paisible** a *un jardinier noir*, paisible, *lui* aussi, et soumis à la fatalité. *Ce jardinier* a <u>un fils</u> et <u>ce fils</u> participe à une manifestation d'écoliers violemment réprimée par la police. On apprendra que <u>l'enfant</u> est arrêté, puis qu'<u>il</u> est mort. (Résumé d'*Une saison blanche et sèche*, *Télérama*, 15.01.92)

Dans cet exemple, la chaîne référant au professeur (en gras) s'achève quand débute celle du jardinier (en italique). On peut ajouter à l'analyse une troisième chaîne puisque celle du fils (souligné) débute à son tour quand s'interrompt celle du jardinier.

Lorsque deux ou plusieurs référents sont instanciés d'abord séparément puis simultanément dans la même phrase, avec modification de leur statut syntaxique, Schnedecker (*ibid*.) parle de *relations d'entrecroisement.* C'est le cas dans (96) où le personnage de Paulette (en italique) est d'abord mentionné en fonction complément du nom, puis sujet et complément du verbe quand le second référent (souligné) apparaît. L'expression référentielle *Paulette* redevient alors sujet :

> (96) En juin 1940, les parents de la petite *Paulette* (5 ans) sont tués sur une route de l'Exode. *La petite fille* se retrouve seule dans un champ, portant le cadavre de *son* chien. <u>Un jeune paysan</u>, <u>Michel Dollé</u> (11 ans), *l'*amène à la ferme de <u>ses</u> parents, qui décident de *la* garder. *Paulette* veut faire comme les adultes et donner une sépulture à *son* chien. Pour *elle*, <u>Michel</u> organise un cimetière d'animaux. (Résumé de *Jeux Interdits*, *Télérama*, 26.08.92)

Dans les deux types de relations précédents, un référent finit par s'éclipser et laisser la place à un autre. Or, il existe des relations entre chaînes dans lesquelles ce n'est pas le cas, comme le fait remarquer Schnedecker (1997 : 14) :

> « les chaînes peuvent s'interrompre sans que pour autant leur référent ne disparaisse de la circulation référentielle. Il y a en effet des cas où un maillon peut générer un véritable réseau référentiel moyennant des procédés plus ou moins diversifiés et complexes ».

Ces procédés sont caractérisés par le fait que l'introduction d'un nouveau référent ne procède pas par le biais d'un SN indéfini ou d'un nom propre mais peut dériver d'une autre chaîne par des moyens anaphoriques (Schnedecker & Landragin, 2014). Ils se réalisent par l'intermédiaire de relations anaphoriques coréférentielles, telles que les anaphores fidèles ou infidèles comme les anaphores possessives, par exemple dans (97) où le SN possessif introduit le référent repris dans la phrase d'après :

> (97) Capitaine Renard allait de compagnie
>
> Avec son *ami Bouc des plus hauts encornés.*
>
> *Celui-ci* ne voyait pas plus loin que *son* nez. (La Fontaine, III, 5)

Il peut s'agir également de procédés se réalisant par le biais des anaphores associatives. Ainsi dans (98) :

> (98) Jack est ingénieur du son. Un soir, alors qu'il se promène près d'une rivière pour enregistrer les bruits de la nature, il assiste à un *accident*. Il sauve la *passagère* mais *l'homme au volant* est tué (…). (Résumé de *Blow Out*, *Télérama*, décembre 91)

Bien qu'il n'ait pas été précisé de quel type d'accident il s'agissait, le lecteur comprend que ce devait être un accident de voiture grâce aux expressions référentielles *l'homme au volant* et *la passagère*, car dans une voiture il y a un volant et possiblement des passagers.

En outre, différents types de mouvement peuvent intervenir entre les chaînes. Il peut s'agir d'une fusion, à travers laquelle des référents présentés isolément sont constitués en un ensemble au sein d'une seule chaîne. C'est le cas par exemple dans (99) où *Joëlle* et *Camille* sont d'abord deux entités distinctes puis deviennent *elles* et *les deux jeunes* filles :

> (99) *Une fille, Joëlle*, vêtue d'une robe de mariée, est battue par un homme qui *l'*abandonne. Passe *Camille Pelvoux*, avec un caddie sur lequel s'accrochent des mouettes. *Elle* pousse *Joëlle* du pied, *lui* déverse des sardines sur la tête, puis *la* charge dans son caddie et *la* ramène chez *elle*. *Elles*

sympathisent et jouent des tours à François […]. *Les deux jeunes filles* se sauvent et arrivent, de nuit, sur le tournage d'un film […]. (Résumé de *Merci la vie*, *Télérama*, 23.08.92)

À l'inverse, il peut s'agir aussi d'une chaîne présentant deux ou plusieurs personnages comme un tout qui se divise en autant de chaînes que d'entités la composant, comme dans (100) où *Deux hommes* devient *l'un* et *l'autre* :

(100) *Deux hommes* parurent. L'*un* venait de la Bastille, l'*autre* du jardin des Plantes. *Le plus grand*, vêtu de toile, marchait le chapeau en arrière, le gilet déboutonné et sa cravate à la main. *Le plus petit*, dont le corps disparaissait dans une redingote marron baissait la tête sous une casquette à visière jaune. (Flaubert, *Bouvard et Pécuchet*)

Ainsi, les chaînes de référence entretiennent entre elles divers types de relations, qui peuvent influer sur la nature de leurs maillons, comme c'est le cas par exemple avec la nécessité de réinstanciation d'un référent suite à l'entrecroisement de deux chaînes.

2.2.3. Les chaînes de référence et les genres textuels

Les chaînes de référence sont caractérisées par la diversité du matériau lexical qui en compose les maillons. Il suffit de revenir aux exemples précédents pour s'en convaincre : un référent peut être identifié par une très grande variété de formes : nom propre, syntagmes nominaux (démonstratifs, possessifs, défini, indéfini etc.), pronoms mais aussi apposition, déterminant zéro, etc. Comme certains travaux l'ont montré, cette variété du matériau lexical peut être corrélée au genre discursif.

Des études montrent que la composition des chaînes de référence varie en fonction du genre textuel. Condamines (2005) compare différents genres allant des descriptions techniques au roman en passant par le texte de loi ou l'article de journal. Elle démontre que les catégories d'anaphores varient d'un texte à l'autre : les anaphores hyperonymiques, par exemple, sont quatre fois plus nombreuses dans les manuels techniques que dans les romans. De la même manière, Longo (2013) étudie des extraits de presse, des textes à caractère purement informatif et un roman en examinant plusieurs critères. Ces critères, empruntés à la typologie de Schnedecker (2005) sont notamment : la longueur des chaînes, la distance inter-maillonnaire, la catégorie grammaticale du premier maillon et des autres... Longo (2013) montre ainsi que la longueur moyenne d'une chaîne dans un roman est plus de deux fois supérieure à celle d'un article de journal. En outre, dans un article de journal, le premier

maillon est généralement un nom propre, alors que le premier maillon dans un roman aura plutôt tendance à être un SN indéfini.

De nombreuses autres études portent soit sur un genre en particulier : le portrait journalistique, le fait divers, les documents techniques ou les discours instructionnels par exemple, soit sur une comparaison de plusieurs genres : presse, roman et textes administratifs, ou encore nouvelles et portraits journalistiques[17]. Cependant, bien que certaines caractéristiques des chaînes de référence se distinguent plus ou moins en fonction du genre discursif, il n'existe pas à l'heure actuelle de typologie qui permettrait de les classer en fonction de leur composition ou de les modéliser.

Une précision s'impose ici quant à l'emploi du terme « genre discursif ». Nous avons remarqué une certaine diversité d'emploi de cette notion dans beaucoup de travaux existants. Les auteurs regroupent sous le terme de « genre » à la fois des genres de discours et des types de textes, sans les distinguer ni les définir précisément. Il est vrai que la notion de genre de discours est une notion très large et sujette à beaucoup de discussions théoriques dans la communauté des linguistes. Comme l'explique Branca-Rosoff (1999 : 6), le genre peut se décrire selon différents points de vue. D'un point de vue formel, la notion de genre peut s'appréhender à l'aide d'une définition socio-historique. Cette approche repose encore aujourd'hui sur « une typologie de l'espace social », comme en témoigne l'étude des discours politiques. Un autre courant de la linguistique s'appuie sur les modes d'énonciation pour classer les discours à l'aide d'éléments qui définissent la situation d'énonciation : le temps, le lieu, les diverses modalités (Bronckart & *al.*, 1985). Certains chercheurs s'appuient sur une typologie purement linguistique (Biber, 1988) en dégageant des critères tels que marqueurs de temps, adverbes de temps et de lieu, formes nominales, etc. pour faire émerger différents genres de texte. Sans entrer ici dans le détail[18], nous prendrons comme critère définitoire pour notre propre analyse celui d'Adam (2012). Dans cette perspective, Adam distingue deux types de genres. D'une part, les genres textuels qui regroupent les textes narratifs (dans lequel s'inscrivent les textes de notre corpus), argumentatifs, explicatifs, descriptifs, dialogaux/conversationnels. D'autre part, les genres discursifs, c'est-à-dire les genres propres à une sphère d'activité sociale, tels que l'enseignement, la presse, la politique, la littérature, etc.

[17] Pour une revue de ces travaux, nous renvoyons à Schnedecker & Landragin (2014).

[18] Pour une synthèse des différents travaux, nous renvoyons à Branca-Rosoff (1999) et Paveau (2012)

2.2.4. Les chaînes de référence et le genre narratif

Dans la mesure où les productions d'élèves qui font l'objet de notre étude sont des textes appartenant au genre narratif[19], nous nous intéressons à présent à ce genre en particulier. Le but de cette section est de montrer les procédés des scripteurs experts à travers l'étude de romans notamment, afin de les comparer plus tard à ce que font les apprentis scripteurs de notre corpus.

Dans les textes du genre narratif, on peut rencontrer certains éléments spécifiques sur lesquels s'appuyer pour identifier des caractéristiques de la composition des chaînes de référence les constituant. Givón (1983, 1989), dans des travaux consacrés au genre narratif et plus particulièrement au roman, a pris en compte plusieurs paramètres tels que distance, persistance et interférence pour distinguer des chaînes de référence de topiques « continus » (c'est-à-dire des référents saillants) et « discontinus » (référents non saillants). Ainsi, selon lui, dans le genre narratif la chaîne de référence d'un topique « continu » (soit un référent saillant) devrait être majoritairement composée de pronoms personnels dans au moins dix phrases consécutives. Pour le monter, nous nous appuyons sur un exemple que nous empruntons à Schnedecker & Landragin (2014), (souligné par eux) :

> (101) Comme le leader des étudiants, **le doyen des études** éprouvait le besoin, moral et politique, d'observer ce qui se passait dans l'anti-institution. Pour des raisons fort différentes, en majorité dues à **sa** loyauté et à **son** attachement pour Gerard Wijnnobel, **il** savait qu'**il** devait surveiller ce que faisait et prêchait Eva Wijnnobel. **Il** savait qu'**il** n'était pas taillé pour s'acquitter de tâches d'infiltration ou de confrontation. **Il** était par instinct profondément libéral et partisan du laisser-faire. **Il** avait accepté le poste de doyen parce que, d'une part, tout le monde était obligé d'assumer à tour de rôle une fonction d'autorité et, d'autre part, **il** voulait rendre les choses plus faciles et plus libres pour les étudiants. **Il** savait mieux **s**'y prendre avec des réactions instinctives de nervosité dans **son** propre camp – restrictions, répressions, exclusions – qu'avec une opposition doctrinaire qui s'opposait pour le plaisir de s'opposer, ce qu'**il** percevait mais **Ø** n'arrivait absolument pas à comprendre. (A.S. Byatt, *Une femme qui siffle*, 2003)

[19] La consigne de la tâche-problème proposée est de raconter une histoire dans laquelle les élèves doivent insérer trois phrases imposées (cf. partie II méthodologie et mise en œuvre de l'étude).

Dans cet exemple de texte narratif, la première mention du référent s'effectue à l'aide du SN défini *le doyen des études.* Dans toutes les autres phrases, le référent est repris sous la forme du pronom anaphorique de troisième personne *il*.

Les chaînes de référence d'un topique discontinu (ou référent non saillant) devraient, quant à elles, être plus courtes et instancier davantage de SN que de pronoms dans la mesure où elles entrent en compétition avec d'autres chaînes (Schnedecker et Landragin, 2014).

Dans leur étude comparative de cinq genres textuels (articles politiques, éditoriaux, romans, textes de lois, rapports publics), Longo et Todirascu (2009) montrent que les chaînes de référence sont plus longues lorsqu'il s'agit de roman que de tout autre genre de texte. En moyenne, une chaîne comporte alors une dizaine de maillons.

En ce qui concerne la distance moyenne entre chaque maillon d'une même chaîne, cette étude montre que, pour certains genres comme le roman, tous les maillons de la chaîne sont localisés dans la même phrase ou dans la phrase suivante. Schnedecker & Longo, dans leur étude de (2012), font le même constat.

Dans un genre narratif tel que le roman, il s'avère que les premiers maillons des chaînes de référence sont principalement des SN indéfinis (Corblin, 1985 ; Longo et Todirascu, 2009 ; Longo, 2013). Toutefois, Landragin & *al.* (2015 : 11) précisent qu'une chaîne coréférentielle prototypique d'un référent (humain) peut débuter par un nom propre et se retrouver sous la forme : nom propre – description définie – succession de pronoms personnels de troisième personne.

L'analyse de la nature la plus fréquente des maillons des chaînes de référence dans le roman (Longo & Todirascu, 2009 ; Schnedecker & Longo, 2012, Landragin, Tanguy, Charolles, 2015) révèle qu'en moyenne, les expressions référentielles dans ce genre de texte ont plutôt tendance à être des pronoms et des SN possessifs avant d'être des SN définis puis des SN indéfinis. L'usage dominant du pronom, à partir de la deuxième mention, n'est pas une surprise puisqu'il était déjà admis que cette forme est la plus couramment employée dans les chaînes de référence (Combettes, 1986 ; Schneuwly, 1988), « dans des rapports qui vont du tiers à la moitié des expressions référentielles totales » (Masseron & Schnedecker, 1995 : 4). Le pronom est la forme « non marquée » de la reprise (Kleiber, 1994), donc la plus facilement et longtemps utilisable, jusqu'à la nécessité d'une redénomination.

La redénomination d'un référent est l'indice d'un problème référentiel (Masseron & Schnedecker, 1995) qui peut être de différentes natures. La première cause de nécessité de redénomination (ou premier type de problème référentiel) tient à la distance inter-maillonnaire qui sépare deux mentions référentielles. Pour illustrer cela, nous empruntons un exemple à Jules Verne (102) :

> (102) Le 29 mai de cette année-là, *un berger* surveillait son troupeau à la lisière d'un plateau verdoyant, au pied du Retyezat, qui domine une vallée fertile, boisée d'arbres à tiges droites, enrichie de belles cultures. Ce plateau élevé, découvert, sans abri, les galernes, qui sont les vents de nord-ouest, le rasent pendant l'hiver comme avec un rasoir de barbier. On dit alors, dans le pays, qu'il se fait la barbe – et parfois de très près. *Ce berger* n'avait rien d'arcadien dans son accoutrement, ni de bucolique dans son attitude. […] (Jules Verne, *Le château des Carpathes,* notre exemple).

Dans cet exemple, l'intervalle entre les deux mentions du berger est trop grand (plusieurs lignes) et les sujets abordés trop variés (paysage, vents, etc.) pour que le référent puisse être récupéré par le lecteur sous la forme d'un pronom. La redénomination vient ici réactiver le référent en question.

La redénomination s'impose également lorsqu'il y a un risque d'ambiguïté référentielle entre deux entités, tel que l'illustre (103), issu d'un texte d'élève et emprunté à Masseron & Schnedecker (1995) :

> (103) *L'homme* arriva dans une station-service et demanda au *pompiste* de l'essence. *Le pompiste lui* a fait le plein et *lui* demanda de l'argent. *L'homme* énervé tira sur *le pompiste* et quitta la station à fond.

Ici, le premier référent *l'homme* se voit concurrencé par l'introduction d'un deuxième référent *le pompiste* avec les mêmes caractéristiques morphologiques. L'emploi d'un pronom ne permettrait pas de déterminer lequel des deux fait l'action.

Ainsi, dans le genre narratif, les chaînes de référence ont quatre caractéristiques principales :

1) elles se caractérisent par la nature des maillons qui les constituent : elles sont composées d'une majorité de pronoms pour les référents saillants et une majorité de SN pour les référents non saillants. Le premier maillon qui introduit un référent est principalement un SN indéfini ou un nom propre.

2) elles se caractérisent par la concentration de leurs maillons dans la même phrase ou la phrase suivante.

3) leur longueur est plus importante que celle des chaînes que l'on trouve dans d'autres genres discursifs

4) Dans le cas où la première mention se trouve trop éloignée de sa reprise anaphorique, ou qu'il y a un risque d'ambiguïté référentielle, la redénomination s'impose.

2.2.5. Les chaînes de référence et la construction des personnages dans les textes narratifs

Dans un récit, les personnages occupent une place centrale. Leur introduction et leur maintien tout au long du texte ne peuvent se faire de manière aléatoire, sous peine de perdre le lecteur. Mais ces dénominations et redénominations peuvent aussi être l'objet d'enjeux stylistiques et ne produisent pas les mêmes effets selon les expressions référentielles employées. Charolles notamment, s'est intéressé à la construction des personnages dans ses travaux, par exemple à travers l'étude récente de *L'Occupation des sols*, d'Echenoz (Charolles & Storme, 2015 ; Landragin, Tanguy, Charolles, 2015). Ce court roman met en scène un père (désigné par son nom de famille, *Fabre*) et son fils (*Paul*) dont la mère (*Sylvie*), défunte, avait fait l'objet d'une publicité avant sa mort. Son effigie est donc placardée sur le mur d'un immeuble voué à disparaître, symbolisant le travail de deuil des deux autres personnages.

Charolles & Storme (2015) analysent l'introduction et la reprise des personnages, à savoir le père, le fils et la mère défunte, ainsi que son effigie (et quelques personnages secondaires). Les trois personnages centraux sont introduits dès la première phrase de l'*incipit,* que nous reproduisons ici sous (104) :

> (104) Comme tout avait brûlé – la mère, les meubles et les photographies de la mère -, pour Fabre et le fils Paul c'était tout de suite beaucoup d'ouvrage : toute cette cendre et ce deuil, déménager, courir se refaire dans les grandes surfaces.

L'introduction du personnage de la mère se faisant au moyen d'un syntagme nominal défini non saturé (*la mère*), il n'est pas possible de savoir de qui elle est génitrice. Ce personnage est introduit ici : « en fonction du rôle dans [l'univers de discours] et non du lien qu'elle entretiendrait avec un de ses enfants » (Charolles & Storme, 2015 : 3).

Tout au long du texte, Jean Echenoz joue avec les redénominations du personnage féminin : en référant à l'aide du nom propre *Sylvie* ou *Sylvie Fabre* à la fois au personnage de la mère et à son effigie, il jette la confusion entre les deux, comme l'expliquent Landragin, Tanguy, Charolles (2015 : 7) :

> « C'est une première spécificité de *L'Occupation des sols* : la mère et son effigie occupent une place centrale, et, si l'on est bien en présence de deux référents, les expressions référentielles utilisées pour l'un ou l'autre sont quasiment identiques, incitant à ne pas les distinguer. »

La chaîne de référence concernant la mère débute de manière inhabituelle et se déroule de manière à entretenir une certaine confusion entre la mère et son effigie. En revanche, il n'en va pas de même pour les personnages masculins. Ceux-ci sont introduits de façon classique par le nom de famille pour le père, *Fabre,* et un SN défini pour le fils, *le fils Paul.* Dans la suite du texte, ils font l'objet de redénominations principalement à l'aide de pronoms. Seul le fils est parfois nommé par son prénom, *Paul,* alors que le père reste dénommé *Fabre*. Ce procédé produit peu de proximité psychologique avec les personnages, comme le font remarquer Landragin & *al.* (2015). Vers la fin du texte, les deux personnages sont amalgamés en un groupe marqué par l'emploi du pronom de troisième personne *on*, lui-même mélangé à d'autres *on* qui incluent petit à petit le lecteur et produisent un effet d'effacement des personnages.

Ainsi, l'analyse particulière de la construction des personnages d'un texte narratif littéraire montre qu'il est possible de trouver des indices concernant les personnages, à travers l'examen de la nature des formes de référence et la manière dont débutent et se déroulent les chaînes de référence de chacun des personnages. De surcroît, le choix des expressions référentielles et leur utilisation dans la chaîne de référence peut influer sur l'effet produit sur le lecteur et relève d'une étude stylistique du texte.

SYNTHESE des notions centrales du chapitre 2

Au terme de ce chapitre, nous pouvons dire que l'anaphore est un phénomène complexe qui a donné lieu à de nombreux travaux. Jusque dans les années 80 environ, la conception « textuelle » de l'anaphore faisait l'unanimité. Pour celle-ci, le contexte linguistique suffit à localiser le référent d'une expression anaphorique. Or, cette conception de

l'anaphore ne montre en rien comment accéder à la « bonne » source. Dès lors, de nouvelles analyses ont été proposées, avec des conceptions et des approches diverses et discutées quant à la définition même de l'anaphore, mais aussi à la conception des processus d'interprétation référentielle et aux statuts des mécanismes d'interprétation. De nouvelles analyses ont aussi été proposées à propos de la description des données anaphoriques elles-mêmes (Reichler-Béguelin, 1988a, 1989 ; Kleiber, 1994 ; Apothéloz, 1995a). La tendance actuelle qui se dégage pour l'analyse de l'anaphore, et dans laquelle nous nous inscrivons, est de mieux prendre en compte les processus d'interprétation référentielle pour montrer que les calculs inférentiels sont autant d'aides à l'identification des référents. Les seules données linguistiques sont importantes mais ne suffisent pas, comme le précise Kleiber (1994 : 11) :

> « L'idée fondamentale est que les seules données linguistiques ne suffisent pas à atteindre le référent visé. […] D'autres facteurs, essentiellement des facteurs cognitifs (accessibilité, pertinence, etc.) entrent en ligne de compte, qui militent pour des approches « moins » linguistiques. »

Au-delà de l'anaphore qui fonctionne avec deux mentions, nous avons vu que les chaînes de référence recouvraient toutes les mentions d'une entité, tout au long du discours. Moins sujet à discussion que la notion d'anaphore, le concept de chaîne de référence a d'abord été étudié, jusque dans les années 2000 environ, sous un angle définitoire. Ces études, notamment les travaux de Schnedecker (1997), ont acté la nature des chaînes de référence : elles débutent à partir de trois maillons (ou trois expressions coréférentielles). En outre, la distance entre les maillons a un impact sur le sentiment de l'interlocuteur quant à la substance (*texture*) de la chaîne. Par ailleurs, la nature des maillons est un facteur de bornage de la chaîne (le nom propre en débute une nouvelle) et les relations qui unissent les diverses chaînes tout au long du discours sont de natures différentes (succession/entrecroisement, dérivation, fusion/partition). À partir des années 2000, les études ont plutôt porté sur l'impact des différents genres textuels sur les chaînes de référence. Bien qu'il n'existe pas (encore ?) de typologie établie des chaînes de référence en fonction des genres textuels, nous pouvons remarquer que dans le genre narratif, qui est celui qui nous intéresse au premier chef pour l'analyse de notre corpus, les chaînes de référence sont souvent modélisées comme suit :

$$\left\{\begin{matrix}\text{SN indéfini}\\ \text{Nom propre}\end{matrix}\right\} - \text{description définie} - \text{succession de pronoms personnels}$$

Les travaux que nous venons d'analyser portent sur des productions d'experts. Aussi, il est intéressant d'envisager la possibilité de les comparer aux productions d'apprenants. C'est ce que nous nous proposons de faire dans le chapitre suivant.

CHAPITRE 3 : Aspect développemental : acquisition des marques anaphoriques dans les productions orales et écrites d'enfants en fonction de leur âge

Les marques de cohésion textuelle relatives à la référence ont beaucoup été étudiées d'un point de vue linguistique, comme nous l'avons vu dans le chapitre précédent. Dans la mesure où notre propre étude s'intéresse aux textes d'élèves, dans une approche à la croisée de la linguistique, de la psycholinguistique et de la didactique, nous ne saurions faire l'économie d'une étude développementale. En effet, l'ensemble des études sur le développement de l'organisation discursive observe une évolution progressive dans la maîtrise des marques de cohésion textuelle. Aussi, en 3.1., nous décrivons les différentes phases de l'acquisition des marques de la référence chez les enfants, que ce soit pour introduire un référent ou le maintenir tout au long du discours. En 3.2, nous mettons en évidence quels sont les facteurs qui influent sur le choix des marques de référence, en fonction de l'âge des enfants, en explorant des études portant sur des productions orales et écrites. En 3.3., nous terminons en nous intéressant à la place de la gestion de la cohésion textuelle dans les processus rédactionnels.

3.1. Les marques anaphoriques en compréhension et en production

L'acquisition des marques anaphoriques dans une perspective développementale a fait l'objet de beaucoup de travaux entre les années 70 et la fin des années 90. Ils sont nettement moins nombreux depuis, les recherches s'étant diversifiées et portant sur d'autres types de constructions spécifiques[20]. En outre, les premières recherches ont beaucoup plus étudié les phénomènes anaphoriques du point de vue de la compréhension que du point de vue de la production. De plus, dans les travaux s'intéressant à la production, de très nombreuses études

[20] Nous renvoyons à Berman (2014) pour un panorama de la question

concernent les enfants entre 0 et 3 ans. Au-delà de cette tranche d'âge, les travaux existants sont moins nombreux.

Les travaux qui adoptent une perspective développementale permettent de donner des éléments de réponse aux deux questions suivantes :

1) Quelles marques anaphoriques les enfants utilisent-ils pour introduire et maintenir les référents dans un discours ?

2) A quel âge ces marques apparaissent-elles dans les différentes phases du développement ?

Pour répondre à ces questions, nous appuyons notre synthèse sur les travaux de De Weck (1991), qui s'est particulièrement intéressée à l'acquisition du langage et notamment à l'apprentissage de la production orale et écrite des discours chez les enfants. Nous appuyons également notre synthèse sur les travaux de Berman (2004, 2008) qui portent eux aussi sur l'acquisition du langage mais dans son développement tardif, autrement dit à l'adolescence ou à l'âge adulte, ainsi que dans plusieurs langues.

Les études développementales sur l'anaphore se sont intéressées à l'emploi des marques référentielles du point de vue de la compréhension. Comme l'explique De Weck (1991), les recherches visent à savoir comment sont compris les pronoms et les syntagmes nominaux (en particulier les déterminants du nom) à valeur anaphorique dans des phrases. Ces travaux mettent en évidence un certain nombre de stratégies de compréhension et de facteurs linguistiques influençant cette compréhension. Bien que la compréhension des phénomènes anaphoriques ne soit pas notre sujet d'étude, nous avons choisi d'y consacrer une partie de cette présentation car les données et les théories issues de ces recherches ont servi de base aux réflexions relatives à la production des anaphores.

Nous l'avons vu, l'anaphore a été moins étudiée du point de vue de la production que de celui de la compréhension. Toutefois, les études développementales sur la production des anaphores montrent les âges auxquels les unités apparaissent dans le langage des enfants et le contexte dans lequel elles sont employées. En 3.1.1., nous décrivons comment les enfants comprennent les phénomènes d'anaphore. Puis, en 3.1.2., nous mettons ces éléments en regard de la production des marques anaphoriques.

3.1.1. Le traitement des marques anaphoriques en compréhension

Comme le rappelle De Weck (1991), les études psycholinguistiques menées du point de vue de la compréhension ont porté sur différents pronoms (personnels et relatifs principalement) et déterminants (définis et indéfinis), mais aussi sur diverses fonctions (exophore, générique, cataphore, anaphore)[21]. Dans la mesure où notre recherche porte sur la référence et l'anaphore dans les textes narratifs écrits par des enfants, nous nous appuyons sur les résultats des travaux expérimentaux concernant la compréhension des pronoms personnels et des déterminants par les enfants.

Les pronoms personnels en compréhension

Les recherches psycholinguistiques qui étudient la compréhension des marques de la référence que sont les pronoms s'intéressent particulièrement aux pronoms personnels de troisième personne (*il/elle* et *le/la/lui*). Plus précisément, il s'agit de comparer la compréhension des pronoms selon différentes fonctions référentielles (anaphore, cataphore, générique…) afin de déterminer laquelle (ou lesquelles) est comprise avant les autres. Par exemple, Kail & Léveillé (1977) demandent à 5 groupes de 20 enfants français de 6 à 14 ans (CP, CE1, CM1, 4° et 5°) d'écouter des phrases lues par un enseignant. Ces phrases sont choisies selon le caractère d'ambiguïté de la référence du pronom. Elles peuvent être :

- sans ambiguïté :

  (105) Quand il est arrivé dans la classe, la maîtresse a salué le directeur.

- ambiguës :

  (106) Hélène a embrassé Marie quand elle est entrée dans la pièce.

  (107) Dès qu'il est monté dans le train, Stéphane a vu Jacques.

À la suite de la lecture des phrases, les enfants répondent à des questions à propos de la phrase qu'ils ont entendue, telles que pour (106) : « Qui est entré dans la pièce ? » ou « Qui a embrassé Marie ? ». Les résultats de cette expérience montrent que les pronoms anaphoriques

[21] Pour un rappel des notions d'*exophore, générique, anaphore et cataphore*, nous renvoyons au chapitre 2, section 2.1.1.1.

sont compris plus précocement que les pronoms exophoriques, les plus jeunes enfants ayant tendance à rechercher le référent à l'intérieur de la proposition.

D'autres travaux arrivent aux mêmes conclusions, comme par exemple ceux de Umstead & Leonard (1983). Ces chercheuses considèrent 3 groupes d'âges différents (3 ans, 4 ans et 5 ans) composés chacun de 10 enfants dont la langue maternelle est l'anglais. Après avoir lu différentes petites histoires de 8 phrases chacune, elles demandent aux enfants à quel personnage le pronom personnel *he* fait référence dans chaque histoire. Cette étude met en évidence le fait que jusqu'à l'âge de 5-7 ans, les enfants comprennent mieux la valeur anaphorique des pronoms, dans des phrases comme dans un texte entier. L'erreur d'identification la plus fréquente, notamment autour de 4 ans, consiste en une surgénéralisation des cas les plus familiers, c'est-à-dire l'anaphore : l'enfant choisit comme coréférent du pronom le personnage dont le nom précède immédiatement le pronom (stratégie de distance minimale, cf. plus loin). Cette erreur est due à une stratégie d'identification de la référence erronée. Kail & Léveillé (1977) expliquent que pour réussir à identifier à quel personnage les pronoms réfèrent, les enfants peuvent utiliser plusieurs stratégies. La première, que les chercheuses appellent *stratégie lexicale*, consiste au repérage des marques de genre et de nombre du pronom pour déterminer le SN auquel le pronom réfère : *il* réfère à un SN masculin singulier par exemple. Toutefois, dans certaines phrases ambigües telles que dans (106), on ne peut pas s'appuyer sur la présence de ces marques pour déterminer le SN de référence. Une autre stratégie peut être utilisée par les sujets : *la stratégie de non-changement de rôle*. Cette stratégie de compréhension consiste à conserver dans la deuxième proposition (celle contenant le pronom) les relations fonctionnelles de la première proposition, c'est-à-dire notamment les positions syntaxiques. Par exemple, dans l'exemple (106) *supra*, le pronom *elle* réfèrerait à *Hélène*. Une troisième stratégie consiste à prendre pour référent le SN le plus proche du pronom (*stratégie de proximité ou de distance minimale)*. De cette manière, dans (106), le pronom *elle* réfèrerait alors à *Marie*. C'est une surgénéralisation de cette dernière stratégie qui est la cause de l'erreur d'identification la plus fréquente chez les enfants les plus jeunes.

L'étude de Kail & Léveillé (1977) montre également que « la compréhension de la coréférence des pronoms personnels par l'enfant renvoie à des stratégies mettant en jeu des niveaux distincts » (*ibid*. : 91). En effet, la stratégie lexicale apparaît plutôt tardivement : à 9;9 ans, quand le pronom précède le SN, les chercheuses constatent encore des transgressions lexicales comme l'attribution de la référence à un agent extérieur à la phrase.

En outre, les chercheuses mettent en évidence le fait que l'ambiguïté de la référence du pronom n'est détectée spontanément qu'à 14;5 ans, ce qui suppose l'existence de réelles capacités réflexives sur la langue à partir de cet âge puisque les sujets doivent envisager qu'un même item linguistique puisse renvoyer à des réalités extralinguistiques distinctes.

Ces travaux montrent ainsi un problème d'attribution de signification. Lorsque des enfants sont confrontés à un ensemble de phrases comprenant des pronoms qui sont exophoriques, anaphoriques ou cataphoriques, ils cherchent prioritairement « un autre élément linguistique coréférentiel pour leur attribuer une signification » (De Weck, 1991 : 65). Et ceci d'autant plus s'il y a « compatibilité des marques de genre du pronom et des SN de la phrase » (Kail & Léveillé, 1977 : 91).

Les pronoms personnels réfléchis en compréhension

L'étude de la compréhension des anaphores pronominales s'est également intéressée aux pronoms personnels réfléchis. Par exemple, Deutsch, Koster et Koster (1986) proposent à 96 enfants néerlandais de 6 à 10 ans d'écouter deux phrases dans lesquelles le pronom personnel de la troisième personne et son antécédent peuvent être réflexifs ou non :

(108) The brother of Piet washes himself

(109) The brother of Piet washes him.

Les expérimentateurs fournissent ensuite des images correspondant à plusieurs interprétations possibles des phrases entendues. Pour chaque phrase, les enfants doivent choisir l'image qui correspond à leur compréhension de la phrase. Cette expérience met en évidence le fait que les pronoms personnels réfléchis semblent maîtrisés plus rapidement que les autres types de pronom.

Les déterminants définis en compréhension

Les pronoms ne sont pas les seules unités linguistiques capables de référer. Des recherches psycholinguistiques s'attachent à décrire la compréhension des syntagmes nominaux définis. Karmiloff-Smith (1979 notamment), dont les nombreuses recherches portent à la fois sur la compréhension et la production des anaphores en anglais, décrit différentes phases dans l'acquisition des principales fonctions des déterminants définis. Selon elle, vers l'âge de 3 ans, les enfants comprennent la fonction déictique. Autour de l'âge de 6

ans, ils comprennent la fonction exophorique. La fonction anaphorique, quant à elle, n'est comprise que vers l'âge de 8-9 ans.

Si l'ordre des acquisitions semble rester le même, l'âge auquel sont acquises ces notions peut varier sensiblement en fonction du type de tâche et d'énoncé. En effet, De Weck (1991) s'appuie sur Karmiloff-Smith (1979 : 191) pour rappeler par exemple que la fonction exophorique des déterminants définis est comprise vers 8 ans quand il s'agit de répondre à des questions à propos d'une courte histoire, dont les référents ne sont pas présents dans le contexte extralinguistique :

> « It is not until 8 years, […] that children pick up the linguistic clue. Before this, children use « pragmatic procedures » i.e. they anticipate the details of a story from their world knowledge of similar situations [22]».

De manière générale, l'acquisition des différentes fonctions de la référence suit la même évolution du point de vue de la production (Karmiloff-Smith, 1979 ; De Weck, 1991), comme nous allons le montrer à présent.

3.1.2. Les marques anaphoriques en production

Les travaux menés sur l'acquisition des marqueurs anaphoriques du point de vue de la production ont pour objectif d'apporter des indications relatives à l'âge auquel les différentes unités et structures de la langue apparaissent dans les productions orales ou écrites des enfants.

De très nombreux travaux se sont intéressés aux enfants entre 0 et 3 ans, dont la langue première est notamment l'anglais ou le français. Ces études montrent que les principales formes des pronoms apparaissent vers l'âge de 2-3 ans (Huxley, 1970 ; Brown, 1973), corroborées par des études plus récentes comme celle de Salazar Orvig & *al*., 2004). À cet âge, les pronoms sont construits en opposition au niveau de la personne (*je*/*il)*, du nombre (*je/nous)*, du genre (*il/elle)* et du cas (*je/me).* Comme l'explique Waryas (1973, cité par De Weck, 1991), les différents niveaux d'opposition résultent de la nature de la situation de

[22] « Ce n'est pas avant l'âge de 8 ans, (…] que les enfants saisissent l'indice linguistique. Avant cela, les enfants utilisent « des procédures pragmatiques » par exemple, ils anticipent les détails d'une histoire en fonction de leur connaissance du monde dans des situations similaires » (notre traduction)

communication, de la nature du référent dans le domaine extra-linguistique, ou des restrictions syntaxiques de la langue elle-même.

L'emploi des déterminants par les jeunes enfants a fait l'objet d'une étude longitudinale menée par Brown (1973). Pendant 5 ans, il a rencontré tous les mois 3 enfants américains (18 et 27 mois), durant 2 heures, pour étudier leur production spontanée du langage lors de conversations quotidiennes. Son étude montre que l'emploi des déterminants apparaît vers l'âge de 3 ans, notamment celui des articles définis et indéfinis. Le chercheur met également en évidence le fait que l'article défini, chez les enfants de cet âge, est employé avec un objectif de référence spécifique, c'est-à-dire pour une entité connue par lui et son interlocuteur dans la situation de communication. L'article indéfini, quant à lui, est employé en référence non spécifique. À partir de 3 ans, les enfants sont donc en mesure d'utiliser dans leurs énoncés des pronoms et des déterminants. Mais l'on peut se demander si ces emplois sont anaphoriques ou exophoriques et à partir de quel âge les enfants sont capables d'emplois anaphoriques.

En réponse à cette question, Huxley (1970) a montré que lorsqu'il s'agit des pronoms, les enfants emploient d'abord le pronom de troisième personne pour référer à quelqu'un « en vue » (déictique) ou connu de l'interlocuteur (fonction exophorique). Plus tard, ils emploient le pronom de troisième personne pour référer à quelqu'un qu'ils ont mentionné dans leur discours (fonction anaphorique). Le même ordre d'acquisition est observé pour les déterminants. Les déterminants sont d'abord employés en fonction exophorique avant de l'être en fonction anaphorique, comme c'est le cas, nous venons de le voir, en compréhension. Cependant, nous pouvons nous interroger sur l'âge auquel la maîtrise de ces fonctions apparaît. Dans les années 70, début des recherches psycholinguistiques sur cette question, un courant de pensée postulait que cette maîtrise est avérée chez les jeunes enfants entre 0 et 3 ans (Huxley, 1970 ; Brown, 1973 ; Tanz, 1974). Aujourd'hui, la recherche s'appuie sur les résultats de nombreux travaux pour affirmer que cette maîtrise est progressive et plus tardive.

Lorsque nous nous intéressons plus particulièrement aux études sur la fonction anaphorique, nous constatons que cette progressivité existe également dans l'emploi des marques de la référence. Dans les sections qui suivent, nous tentons donc de répondre aux questions suivantes relatives à la fonction anaphorique en production de texte, formulées par De Weck (1991) : comment des enfants d'âges différents introduisent-ils des référents dans un texte, en particulier les protagonistes dans une narration (3.1.2.1.) ? Comment maintiennent-

ils la référence, et plus précisément peut-on dégager des lois qui gouvernent l'usage des pronoms personnels et des déterminants (3.1.2.2.) ? De manière plus générale, existe-t-il des facteurs qui influencent les divers procédés anaphoriques dans les narrations (3.1.2.3.) ?

3.1.2.1. L'introduction des référents dans la narration : une maîtrise tardive du SN indéfini en position initiale

Les travaux dont nous nous apprêtons à présenter la synthèse, bien que portant principalement sur l'oral, ont été confirmés par des travaux portant sur la production écrite. Loin d'opposer l'oral à l'écrit, ces études montrent que les compétences acquises à l'oral peuvent être réutilisées à l'écrit, comme l'expliquent De Weck & Schneuwly (1994 : 467) : « the basic capacity acquired in the oral mode can be used when the subjects write narrative texts[23] ».

Comme le fait remarquer Wigglesworth (1990 : 105), l'aptitude à raconter un récit cohérent et structuré requiert des compétences linguistiques et cognitives considérables : l'information doit être organisée en un tout, les participants nommés et différenciés et les référents linguistiques établis pour les relations anaphoriques futures. Il nous semble que ce dernier point est important car de la bonne introduction du référent découle le bon enchaînement des relations anaphoriques avec ses autres mentions et par là-même, la bonne cohésion de l'ensemble de l'énoncé. Cependant, introduire un référent n'est pas aisé. Jisa (2000 : 592) insiste sur le fait que l'introduction d'un référent nécessite des connaissances pragmatiques sophistiquées sur les connaissances partagées ou non par le narrateur et le destinataire. Comme nous l'avons vu dans le chapitre 2 (cf. section 2.2), la manière la plus ordinaire d'introduire un référent dans un discours est d'utiliser un SN indéfini. Si l'adulte maîtrise (en général) ce processus, ce n'est pas forcément le cas pour les enfants. Chez ces derniers, la construction du système des déterminants est « lente et progressive car les enfants doivent distinguer et coordonner plusieurs fonctions : fonction nominative et référence non-spécifique pour l'indéfini, fonction déictique, exophorique et anaphorique pour le défini » (De Weck, 1991 : 85).

Plusieurs recherches aboutissent à la conclusion que l'introduction d'un référent par un SN indéfini se réalise progressivement et tardivement chez les enfants. Par exemple, dans ses

[23] « la compétence basique acquise dans le mode oral peut être utilisée quand le sujet écrit des textes narratifs. » (notre traduction)

travaux, Warden (1981) propose à 80 enfants entre 5 et 8 ans, dont la langue maternelle est l'anglais, de visionner de très courts films mettant en scène un ou plusieurs personnages puis de reformuler ce qu'ils ont vu. Il montre ainsi que si les enfants emploient majoritairement un SN indéfini pour introduire le référent, ils ne le font pas de manière systématique. D'autres études montrent même qu'il faut attendre l'âge de 9-10 ans pour que les enfants introduisent systématiquement les référents par une expression appropriée (indéfinie à fonction non-spécifique). C'est le cas par exemple dans d'autres travaux de Warden (1976). Le chercheur constitue 4 groupes d'enfants de 3, 5, 7 et 9 ans et établit des paires d'enfants dans chaque groupe. Dans chaque paire, un enfant doit raconter une histoire en s'inspirant de dessins que l'autre enfant ne voit pas. Et *vice versa.* Les résultats montrent que plus les enfants sont jeunes, moins ils recourent à un SN indéfini pour introduire un nouveau référent. Comme le montre De Weck (1991, s'appuyant sur les travaux d'Hickmann), avant 9-10 ans, on constate un nombre important d'introductions déictiques, au moyen de défini ou d'indéfini accompagné d'élément déictique (*with a cat there*, Hickmann, 1987 : 248). Chez les plus jeunes (avant 6 ans), on rencontre parfois des introductions sous forme d'*étiquetage* (Hickman, 1984) plus ou moins explicite : *c'est un X, un X,* seul constituant du premier énoncé. Ces étiquetages témoignent de la fonction nominative de l'article indéfini, en opposition avec la fonction déictique de l'article défini. Comme le fait remarquer De Weck (1991 : 86) : « C'est dans la mesure où ces étiquetages attirent l'attention de l'interlocuteur sur un référent, que l'on peut considérer qu'ils servent indirectement d'introduction ». Toutefois, on ne peut affirmer que l'utilisation des déterminants se fait au hasard chez les enfants, quel que soit leur âge. De Weck (1991) s'appuie sur Warden (1976) pour affirmer que : « l'absence quasi-totale de déterminants indéfinis dès la deuxième mention d'un référent signifie que l'enfant sait qu'un article indéfini ne peut être utilisé une fois qu'un référent a été identifié » (De Weck, 1991 : 87).

Après avoir introduit le référent, l'enfant doit pouvoir le maintenir tout au long de son discours pour que son interlocuteur comprenne le sens de son propos. Nous allons voir dès à présent comment les enfants maintiennent la référence dans un discours.

3.1.2.2. Le maintien des référents dans la narration : une construction progressive avec le pronom pour expression privilégiée

Lorsqu'il s'agit de maintenir la référence à une entité, il semblerait que 7 ans soit un âge charnière. C'est en effet à partir de cet âge que les enfants établissent des liens anaphoriques ou intralinguistiques entre les différentes mentions d'un même élément.

Comme l'explique De Weck (1991), Hickmann (1980, 1984, 1987) montre une évolution dans la manière dont les enfants maintiennent la référence. Avant 7 ans, on observe une prédominance des étiquetages plus ou moins explicites qui ré-introduisent souvent le personnage lorsque l'enfant passe d'une image à l'autre « de telle sorte que les énoncés successifs sont à la fois indépendants les uns des autres et dépendants du contexte non linguistique » (De Weck, 1991 : 87). À partir de 7 ans, les enfants manient les pronoms et les déterminants de telle sorte qu'il existe des relations entre les expressions employées pour maintenir la référence et les référents eux-mêmes. C'est le cas par exemple dans une histoire où le protagoniste a un rôle plus central que les autres. Pour maintenir la référence à ce personnage, les enfants emploient des pronoms explicites et des anaphores-zéro alors qu'ils utilisent plutôt des syntagmes nominaux et quelques pronoms pour référer aux personnages secondaires. Hickmann (1987 : 256, cité par De Weck, 1991) conclut ainsi que :

> « ce n'est qu'après 7 ans que les enfants présupposent les référents par rapport au co-texte et/ou ne les « ré-introduisent » plus d'énoncé en énoncé lorsque c'est clairement possible, notamment lorsqu'ils réfèrent souvent à un même personnage principalement présenté comme agent et/ou sujet d'une image à l'autre ».

Karmiloff-Smith (1980, 1981) a étudié l'acquisition de l'anaphore pronominale chez 170 enfants entre 4 et 9 ans dont la langue maternelle est soit l'anglais soit le français. Dans ces travaux, elle demande aux enfants de produire un récit à partir de 6 images. Les résultats de cette étude, résumés par Bamberg (1986 : 239) et Wigglesworth (1990 : 107), ont montré qu'avant l'âge de 6 ans, les emplois pronominalisés ne sont que déictiques. Ce n'est qu'à partir de l'âge de 6 ans que les enfants commencent à utiliser des pronoms anaphoriques pour maintenir la référence, à travers la stratégie dite du sujet thématique. En effet, les enfants de 6 ans et au-delà réservent l'emploi du pronom personnel placé en début de phrase au sujet thématique, qui est en général le personnage principal de leur histoire. Les autres entités du récit sont dénommées à l'aide de SN définis. Les enfants plus âgés (8/9 ans) emploient également le pronom en position initiale dans la phrase pour référer au sujet thématique, mais ils peuvent aussi placer des personnages secondaires en position initiale, qu'ils décrivent alors à l'aide d'expressions définies. Les résultats de cette étude ont permis à Karmiloff-Smith (1983) de déterminer 3 phases du développement de la référence chez l'enfant, résumés par Wigglesworth (1990 : 107) et De Weck & Schneuwly (1994). La première phase, avant l'âge

de 6 ans, est caractérisée par des pronoms utilisés de manière déictique. La deuxième phase, entre 6 et 8 ans environ, est caractérisée par l'emploi de pronoms pour le sujet thématique, pronoms qui apparaissent de manière prédominante en position initiale dans la phrase. Enfin, la troisième phase (à partir de 9-10 ans environ) est caractérisée par l'emploi de pronoms qui peuvent référer à d'autres éléments que le sujet thématique, par exemple « in reduplicating a noun phrase in initial or last slot of utterances[24] » (De Weck & Schneuwly, 1994 : 466).

Ces phases de l'acquisition de la référence ont été confirmées par Bamberg (1986). Ce chercheur étudie la manière dont les enfants et les adultes maintiennent la référence au cours d'un discours en utilisant des formes linguistiques variées. Pour cela, il propose à des enfants monolingues allemands et un de leurs parents un livre de 24 pages sans texte. Les parents doivent raconter l'histoire correspondant aux images à leur enfant. Quelques jours plus tard, les enfants doivent à leur tour raconter l'histoire, en s'appuyant sur le livre d'images. Bien que les enfants soient influencés par la manière de raconter de leurs parents, ces travaux montrent que la plupart des enfants entre 3 et 4 ans, ainsi que quelques-uns autour de 5-6 ans, emploient principalement un pronom en position initiale dans la phrase lorsqu'il s'agit de référer au sujet thématique. À partir de 9 ans, les enfants utilisent des pronoms en position initiale de phrase pour maintenir la référence, même s'il s'agit d'une entité autre que le sujet thématique. En revanche, ils emploient des SN pour introduire ou changer de référent.

De même, dans une étude plus récente, Jisa (2000, 2004) demande à 3 groupes d'âges différents (5, 7 et 10 ans) de 15 enfants chacun de regarder attentivement un livre d'images mettant en scène les aventures d'un homme et d'un chien. Les enfants doivent ensuite raconter l'histoire à un adulte qui n'a pas connaissance des images. Les résultats de l'expérience confirment la prédominance de l'emploi des pronoms dans le maintien de la référence, quel que soit l'âge des enfants. D'autres recherches arrivent au même constat. Bartlett (1984) par exemple, demande à 160 enfants de langue anglaise de 11 à 13 ans de raconter par écrit les évènements qu'ils ont vus au préalable dessinés sur des images. Dans chaque groupe d'âge, les enfants sont divisés en deux sous-groupes : un constitué d'enfants considérés comme bons dans les tâches d'écriture, c'est-à-dire ayant des compétences au-dessus de la moyenne (*good writters*) et l'autre d'enfants médiocres, c'est-à-dire ayant des compétences en-dessous de la moyenne (*poor writters*). L'étude montre que quel que soit le groupe d'âge ou de niveau, les pronoms sont privilégiés dans le maintien de la référence. De

[24] « en reproduisant un SN en position initiale ou en dernière position dans un énoncé » (notre traduction)

même, les recherches de Hickmann (1987, *opp. cité)* ; Hickmann & *al.*, (1995, voir section 3.1.2.3. de ce chapitre) ; Hickman (2003) et Berman (2004) arrivent aux mêmes conclusions. Ainsi, de nombreuses études montrent que le pronom est l'expression référentielle préférée dans le maintien de la référence.

Quelques travaux se sont intéressés à l'utilisation de l'article défini dans le maintien de la référence. Warden (1976), par exemple, constitue 4 groupes (3 ans, 5 ans, 7 ans et 9 ans) d'une vingtaine d'enfants anglais chacun. Dans chaque groupe, le chercheur propose à un enfant une histoire composée de 3 dessins comprenant 3 référents animés et un inanimé. L'enfant doit alors raconter l'histoire à un autre enfant qui n'a pas accès aux dessins. L'expérience révèle que c'est à partir de 9 ans que les enfants limitent leur utilisation de l'article défini dans le discours narratif à l'article défini anaphorique. À partir de cet âge, ils emploient l'article défini pour un référent en seconde mention. Karmiloff-Smith (1979) arrive à la même conclusion. De même, Jisa (2000)[25] indique que l'article défini est principalement employé pour réintroduire un référent, c'est-à-dire placer à nouveau en position sujet un référent déjà introduit dans le discours, ou en d'autres termes, réactiver le référent devenu peu accessible en mémoire (cf. chapitre 1, section 1.2.).

3.1.2.3. La gestion de la cohésion textuelle en production écrite

Les travaux que nous allons présenter maintenant étudient la gestion de la cohésion textuelle dans des productions écrites d'enfants. Bien que s'intéressant également aux modalités d'introduction et de maintien des référents, ils sont toutefois différents des travaux présentés dans les sections précédentes. En effet, il s'agit de tâches rédactionnelles provoquées.

Le premier à s'être intéressé à la gestion de la cohésion textuelle dans les productions écrites d'enfants, et plus particulièrement aux risques d'ambiguïtés référentielles, est M. Charolles (1988c). Partant du constat que dans tout discours, le rédacteur doit anticiper les problèmes d'interprétation et contrôler les risques de confusion entre personnages, Charolles (1988c : 81) a voulu « apprécier aussi finement que possible si les élèves sont sensibles, dans la rédaction de leur texte, aux risques de confusion entre personnages qui peuvent apparaître à

[25] cf. paragraphe précédent pour le détail du protocole expérimental

la lecture ». Dans cet objectif, il propose à 64 enfants de 10 à 13 ans de rédiger l'histoire correspondant à une courte bande dessinée relatant un naufrage. La consigne est la suivante :

> « En précisant bien ce que font les différents personnages sur chacune des vignettes, racontez ce qui se passe dans la bande dessinée (et seulement ce qui s'y passe), sachant que votre lecteur ne l'aura pas sous les yeux. Les vignettes sont dans l'ordre. »

La bande dessinée comprend deux versions : une version supposée facile dans laquelle les trois personnages sont différenciés (couleur des cheveux, couleur des vêtements) et une version considérée comme plus difficile dans laquelle les protagonistes sont indifférenciés (même dessin de personnage pour chacun des trois). Une moitié des enfants reçoit la bande dessinée dite facile et l'autre moitié, celle dite difficile. Dans les deux cas, l'analyse *a priori* de la tâche permet d'établir « le nombre des formes de désignation ou de reprise qu'un lecteur coopératif ne peut solutionner » (*ibid.* : 81). À la lecture des résultats, il s'avère que les élèves les plus jeunes sont ceux qui ont le nombre d'emplois confus le plus élevé (6,02% pour la version facile et 9,75% pour la version difficile). Mais Charolles tempère ces résultats en faisant remarquer qu'aux cas de confusion, vient s'ajouter les cas *d'évitement*. Il s'agit des enfants qui « font l'impasse » sur certaines actions des personnages, notamment dans des vignettes où ils agissent conjointement.

Dans la mesure où l'un des objectifs de nos propres travaux est de déterminer comment les élèves introduisent les référents dans un texte, cette étude propose une typologie des différentes stratégies d'introduction des référents dans une tâche rédactionnelle. Ces stratégies sont les suivantes (*ibid.* : 84) :

- les dénominations passe-partout telles que *un homme, un monsieur, un enfant, deux garçons*, …
- les noms propres
- les rôles imposés (essentiellement réservés dans le corpus aux deux sauveteurs qui se voient promus *pêcheurs, marins, indiens, amis* et même *parents*)
- les dénominations fonctionnelles établies sur la base des actions représentées dans la BD (*les deux sauveteurs*)
- les dénominations mentionnant un trait descriptif comme par exemple *deux sauveteurs, un blond, un brun*.

Dans le corpus recueilli, les dénominations passe-partout sont les plus fréquentes dans l'ensemble des textes (50%), suivies par le recours à un nom propre (30%), principalement

chez les enfants les plus âgés. Dans une moindre mesure, la stratégie consistant à imposer un rôle aux personnages est davantage suivie par les jeunes enfants (14% au CM1) que par les plus âgés (5% en 6°). Enfin, le recours à des traits fonctionnels ou descriptifs est peu fréquent dans le corpus étudié. Ainsi, pour désigner les personnages, les enfants utilisent le plus souvent des dénominations non spécifiques (dénominations passe-partout), les indices descriptifs étant sous-exploités (*ibid.* : 86).

Ces résultats montrent également trois phénomènes qui peuvent expliquer la mauvaise gestion des personnages par les élèves : la décontextualisation, les anaphores confuses et les problèmes liés à l'introduction des noms propres.

La décontextualisation est une mise à distance du texte par rapport au support d'écriture. Comme l'explique Charolles (1988c : 88), les enfants n'arrivent pas à se détacher du support d'écriture : « malgré la consigne invitant à « raconter », nombre d'entre eux rédigent des textes qui décrivent plus qu'ils ne racontent les évènements représentés dans chaque vignette ». C'est le cas par exemple dans le texte suivant, rapporté par l'auteur (*ibid.* 86) et reproduit ici sous (110):

> (110) 1. L'homme à la mer. 2. Deux pêcheurs le voient. 3. Ils le traînent vers une île. 4. Un des pêcheurs le pose sur la plage pendant que l'autre tire le bateau sur la rive. 5. Après l'homme se réveille.

Cette absence de décontextualisation se traduit également par des intrusions de l'auteur dans le récit, telles que « nous voyons l'homme qui s'appuie la main au cocotier » ou des allusions à la position des personnages dans les vignettes : « le premier homme à droite ». En outre, l'insertion de séquences au discours direct à l'intérieur d'un récit au discours indirect (et *vice versa*) est fréquent chez les élèves les plus jeunes, comme en témoigne l'exemple suivant rapporté par l'auteur (*ibid.* : 89) :

> (111) 2 hommes sont dans une petite barque. Ils me prennent chacun par un bras. Je vois une île où ils se dirigent. Un homme me lâche le bras et prend une rame. Nous arrivons sur l'île. L'homme me sort de l'eau et l'autre approche la barque, quelques cocotiers ont poussé. Un homme s'appuie sur un cocotier et fait la triste tête et puis il le dépose par terre il le font sécher au soleil.

Comme le souligne Charolles (*ibid.*), le passage à la troisième personne suppose une distanciation par rapport à l'expérience. Cette mise à distance ne va pas de soi chez les plus jeunes élèves car ils ont du mal à la fois à se « projeter » dans l'action afin de pouvoir l'interpréter, et à la « contempler » de l'extérieur.

Le deuxième phénomène impliqué dans la mauvaise gestion des personnages est celui des anaphores confuses. Ce type d'anaphore, qui touche aux formes de reprise des référents, peut se traduire de trois manières différentes :

- par des syllepses (cf. aussi chapitre 2, section 2.1.2.2.), comme dans l'exemple suivant où le pronom personnel singulier *il* reprend le canoë alors que l'élève entend moins référer au canoë, qui ne peut pas voir, qu'à ses occupants (*ibid.* : 90) :

  (112) Un homme était dans l'eau il ne savait pas nager. Un canoë passe avec deux hommes à bord, il voit l'homme qui [est] entrain de couler ils le …

- par des distinctions de personnages difficiles, tel que c'est le cas dans (*ibid.* : 90) :

  (113) Un homme se noie, il crie « au secours »
  Deux hommes en bateau le voient
  Ils le prennent
  Un marin le tire par le bras jusqu'à une île un le pose sur l'île. L'autre pose le bateau…

  Dans cet exemple, la distinction entre les sauveteurs est rendue difficile par l'introduction de *un marin*, que l'on a tendance à interpréter comme un nouvel individu.

- par des ruptures de chaînes, qui peuvent se traduire par le passage d'un SN indéfini singulier, *un des*, à un pronom personnel pluriel, *ils*, puis à un pronom personnel singulier, *il*, comme dans l'exemple ci-dessous (*ibid.* : 91) :

  (114) … Un des hommes le prit par la main.
  Ils avaient aperçu au loin une petite île
  Il le traîna pendant que l'autre ramait…

Enfin, le troisième phénomène impliqué dans la mauvaise gestion des personnages est lié à la mauvaise utilisation des noms propres. La stratégie qui consiste à nommer les personnages est une stratégie efficace pour éviter les risques d'ambiguïté référentielle. Mais elle devient problématique lorsque l'introduction des noms propres est soudaine et « en cours de route », comme dans l'exemple reproduit ci-après (*ibid.* : 92) :

(115) C'est l'histoire d'un monsieur qui se noyait dans l'eau. Deux garçons arrivèrent, Patrice et Pierre en barque, ils le repêchèrent. Tous les deux en barque ramenèrent monsieur Dupont…

Sans précautions de la part du rédacteur, le passage d'une dénomination passe-partout, *un monsieur*, à un nom propre, *Monsieur Dupont*, qui est censé référer à ce même personnage, peut semer la confusion chez le lecteur.

D'autre part, les travaux de Millogo (2008) ont pour objectif d'étudier l'évolution des modalités de gestion de la cohésion textuelle en fonction du niveau scolaire des enfants. Plus précisément, ces travaux étudient l'effet de trois variables sur l'emploi de pronoms personnels dans un texte. Ces trois variables sont :

- la distance qui sépare la source de sa reprise : il s'agit du nombre de propositions qui séparent la première apparition du référent de sa reprise dans la phrase à trou. Pour la variable *Distance*, deux modalités sont considérées par l'auteur. Dans la première, une proposition sépare la première apparition de sa reprise (distance de 1, soit D1 = une proposition), dans la seconde, ce sont quatre propositions qui séparent la première apparition de sa reprise (distance de 4, soit D2 = quatre propositions)

- la position thématique ou non de la source : il s'agit de faire varier le thème dans la proposition inductive par une proposition clivée en *C'est…que*. Pour la variable *Position thématique*, deux modalités sont considérées par l'auteur. Dans la première, le sujet grammatical est thématisé (soit T1 = sujet grammatical thématisé : ex. *c'est une petite fille qui cherche un petit garçon*). Dans la seconde, l'objet grammatical est thématisé (soit T2 = objet grammatical thématisé : ex. *C'est un petit garçon qu'une petite fille cherche*)

- le changement de focus du discours : il s'agit de faire varier la direction de l'attention, dans l'une des propositions intermédiaires du texte, sur l'un ou l'autre personnage de la première proposition. Pour la variable *Focus du discours*, deux modalités sont considérées par l'auteur. Dans la première, le focus d'attention est maintenu (soit FM = focus maintenu : ex. *c'est une dame qui promène un bébé dans un parc. Des branches mortes craquent sous ses pieds*.). Dans la seconde, le focus d'attention est rompu (soit FR = focus rompu : ex. *c'est une dame qui promène un bébé dans un parc. Le bébé se réveille.*)

Pour étudier les effets de ces variables, Millogo réunit 58 enfants français de 3 niveaux scolaires différents : CE1 (7 ;4 ans), CM1 (9 ;7 ans) et 6° (11 ;7 ans). Le chercheur propose à chaque enfant un carnet contenant 12 courts textes à trous que les enfants doivent compléter, tel que par exemple :

> (116) C'est un monsieur qui conduit une petite fille à l'école. La voiture est garée sur le parking de l'école. …….. discute avec l'instituteur.

Dans chacun de ces textes, la proposition inductive présente deux personnages de genres différents. Le référent de la reprise (ou l'antécédent d'une éventuelle expression anaphorique) est le sujet grammatical de la proposition inductive. La dernière phrase de chaque texte comporte un blanc à la place du sujet grammatical.

Les textes diffèrent en fonction de chaque situation, correspondant au croisement des deux modalités des trois variables manipulées (Distance en 1 ou 4 propositions, Sujet grammatical thématisé ou non, Focus maintenu ou rompu, soit 2x2x2=8). De cette manière, 8 textes différents sont proposés aux élèves. Par exemple, un premier texte est proposé dans lequel il y a une seule proposition entre le référent et sa reprise. Dans ce texte également, le sujet grammatical est thématisé et le focus maintenu. Il s'agit du texte présenté précédemment (116). Un autre texte est proposé dans lequel il y a cette fois 4 propositions qui séparent le référent de sa reprise. Dans ce deuxième texte, le sujet grammatical n'est pas thématisé et le focus est rompu :

> (117) C'est une petite fille qu'un monsieur conduit à l'école. L'école primaire est éloignée de la maison. La circulation automobile est très importante devant l'école. La cour de l'école est très grande. La petite fille traverse la cour pour aller en classe. **…….** discute avec l'instituteur.

Les résultats de cette étude montrent qu'en fonction de leur âge, les enfants appliquent des stratégies spécifiques de gestion des pronoms personnels anaphoriques excepté en ce qui concerne la distance. En effet, c'est la seule variable pour laquelle « dès 7 ans, les participants choisissent moins de pronoms personnels comme marque de reprise lorsque la distance augmente » (Millogo, 2008 : 7). L'augmentation de la distance séparant l'expression référentielle de sa reprise entraîne une diminution de l'emploi de pronoms personnels, quel que soit l'âge des enfants, car le référent est de moins en moins accessible.

Au sujet de la variable concernant la thématisation, il s'avère que les enfants de 9 ans (CM1) utilisent beaucoup plus de pronoms lorsque le référent est thématisé que lorsqu'il ne l'est pas. La thématisation du référent « rend sa trace plus disponible plus longtemps en

mémoire » (*ibid.* : 7). Cependant, l'emploi de pronoms en fonction de la thématisation du référent est beaucoup moins présent dans les autres groupes d'âge (CE1 et 6°).

En ce qui concerne la variable *focus*, il semblerait qu'elle influe de manière différente sur les stratégies des enfants en fonction de leur âge, alors qu'elle ne semble pas avoir d'influence sur les enfants de 9 ans. Lorsque le focus est rompu, c'est-à-dire lorsque le personnage de la première proposition n'est plus celui qui fait l'action dans la deuxième proposition, à 7 ans les enfants emploient plus de pronoms personnels que lorsque le focus est maintenu (même personnage qui fait l'action dans deux propositions consécutives). Alors que les enfants de 11 ans, à l'instar des adultes, en produisent moins. Millogo (2008 : 7) explique cette différence de traitement par le fait que :

> « si à 7 ans les participants utilisent plus d'anaphores pronominales dans les situations difficiles ([focus] rompu), il est fort probable que ces expressions ne sont pas forcément « anaphoriques », mais traduisent plutôt une expression générique de reprise : ils utilisent les pronoms personnels *pour désigner* un référent ».

Cette étude[26] corrobore le fait que la différenciation entre la fonction anaphorique et les autres types de fonctions intervient relativement tard (autour de 9 ans) dans le développement.

Les résultats globaux de cette étude montrent qu'à 7 ans, les enfants emploient le pronom personnel comme une expression *passe-partout*, non spécifique. A 9 ans, les enfants utilisent peu le pronom personnel mais d'autant plus que le sujet grammatical est thématisé. À partir de 11 ans, les pronoms sont principalement employés lorsque le référent est maintenu dans le focus mais la thématisation du référent n'est pas bien prise en compte. Ainsi, « le fonctionnement discursif des expressions anaphoriques n'est pas encore totalement maîtrisé à 11 ans. » (Millogo, 2008 : 8).

Ces travaux sont à la fois complémentaires des recherches citées ultérieurement (cf. sections 3.1.2.1. et 3.1.2.2.) et ont toutefois un statut particulier. Ils se caractérisent par le fait que, contrairement aux travaux précédemment étudiés, Millogo ne s'appuie pas sur l'expression spontanée des enfants pour étudier la référence car il met en place un protocole spécifique relevant de l'épilinguistique. La tâche expérimentale qu'il propose s'apparente à un exercice scolaire (texte à trous). Par sa nature, les enfants ont déjà été confrontés à ce type de tâche, ils peuvent donc répondre exactement ce qui est attendu d'eux puisque c'est un exercice qu'ils connaissent. En outre, le texte à trous restreint fortement les possibilités de

[26] Les résultats de cette étude sont les mêmes que ceux des travaux antérieurs détaillés dans ce chapitre, section 3.1.2.2.

réponse car il pointe un lieu précis du discours dans lequel il est attendu que l'enfant résolve un problème spécifique, de nature métalinguistique. Cette étude présente pour nous un double intérêt : elle est relativement récente et elle s'apparente à celle que nous menons dans notre thèse.

3.2. Les facteurs influençant les stratégies anaphoriques

La production des marques anaphoriques, que ce soit à l'oral ou à l'écrit, se développe donc sensiblement selon les mêmes étapes. Pourtant, il existe des facteurs qui influencent les stratégies anaphoriques.

3.2.1. Effet du type de texte sur les procédés anaphoriques utilisés

Rutter & Raban (1982) ont analysé des textes libres d'enfants de 6 et 10 ans. Les résultats de leur étude montrent que les enfants les plus jeunes (6 ans) ont plutôt tendance à employer des pronoms personnels pour référer. Les enfants plus âgés (10 ans) ont recours à des stratégies plus variées, telles que différentes formes de syntagmes nominaux ou des anaphores nominales. Comme l'indiquent toutefois De Weck & Schneuwly (1994), l'utilisation de textes libres ne permet pas de définir si un type de texte est plus propice qu'un autre à l'emploi de différentes stratégies.

Pellegrini & *al*. (1984) ont voulu déterminer si le type de texte avait un effet sur les stratégies référentielles des enfants. Dans cet objectif, les chercheurs font produire à 71 enfants de 7, 9 et 11 ans un texte narratif et un texte argumentatif, à l'oral et à l'écrit. Les résultats de cette étude indiquent que les enfants emploient plus d'anaphores nominales lexicales dans les textes narratifs que dans les textes argumentatifs. En outre, du point de vue développemental, l'étude montre que la cohésion grammaticale s'acquiert autour de 9/10 ans.

L'emploi de différents types d'anaphore en fonction du type de texte a été confirmé par De Weck et Schneuwly (1994). Ils ont comparé les procédés anaphoriques employés par des enfants de 10, 12 et 14 ans, dans 4 types de textes différents: argumentatif, explicatif,

narratif et compte-rendu d'une expérience personnelle (typologie des auteurs). D'un point de vue développemental, les résultats mettent en lumière le fait que « intralinguistic units of reference [e.g. anaphoric procedures] are acquired relatively late[27] » (De Weck & Schneuwly, 1994 : 475). Le développement de la maîtrise des marques anaphoriques est inégal et couvre une longue période. En outre, les stratégies de reprises anaphoriques employées par les enfants ne sont pas les mêmes en fonction du type de texte. En effet, lorsque l'on regarde dans le texte narratif, qui est celui auquel s'intéressent nos propres travaux, en proportion, les pronoms personnels anaphoriques et les SN possessifs à la troisième personne sont plus nombreux que dans le texte argumentatif. Les SN démonstratifs et indéfinis sont également moins nombreux que dans le texte explicatif. Le type de texte a donc un effet sur le choix des expressions référentielles, comme le montrent également plusieurs autres recherches (Bronckart & Schneuwly, 1991 ; Berman & Katzenberger, 2004).

3.2.2. Effets de la connaissance partagée sur l'utilisation des expressions référentielles

Des travaux ont cherché à étudier l'effet de la connaissance partagée sur l'utilisation des expressions référentielles. Ces travaux nous intéressent particulièrement dans la mesure où ils permettent d'étudier l'anticipation par le rédacteur de la mauvaise compréhension possible par son futur lecteur.

Dans les travaux que nous nous apprêtons à présenter, la situation de connaissance non partagée a été testée à l'oral. Mais cette situation s'apparente également à la production d'écrit. En effet lors de la lecture d'un texte, le lecteur/récepteur ne peut s'aider d'un quelconque appui visuel pour décoder le message et l'encodeur/scripteur ne peut miser sur le fait que le lecteur partage les mêmes informations que lui. L'encodeur/scripteur fera donc en sorte de donner au lecteur toutes les informations nécessaires à la construction de la référence.

Hickmann, Kail et Roland (1995) cherchent à étudier comment la connaissance partagée affecte l'évolution de l'utilisation des expressions référentielles par des enfants français de 6, 9 et 11 ans[28]. À cette fin, il est proposé aux enfants (20 par groupe d'âge) 2 situations. Dans la première, concernant la moitié des enfants de chaque groupe d'âge (soit 10

[27] « Les unités intralinguistiques de la référence [par exemple, les procédés anaphoriques] sont acquises relativement tard » (notre traduction)
[28] voir aussi Bamberg, 1986 ; Wigglesworth, 1990 ; Kail & Hickmann, 1992 ; Kail & Sanchez y Lopez, 1997.

enfants), un livre d'images leur est présenté, relatant des évènements concernant 3 personnages principaux (un garçon, un chien, une grenouille) et 4 personnages secondaires (des abeilles, une taupe, un cerf, un hibou). Chaque enfant doit raconter oralement l'histoire à un expérimentateur adulte qui regarde le livre avec lui (situation de connaissance partagée). Dans la seconde situation (concernant les 10 autres enfants par groupe d'âge), chaque enfant doit raconter oralement l'histoire en s'appuyant sur le livre d'images mais cette fois-ci, à un adulte dont il a au préalable bandé les yeux et qui doit la raconter ensuite à son tour (situation de connaissance non partagée). L'étude révèle que les enfants de 9 et 11 ans utilisent plus d'expressions référentielles (SN indéfinis, pronoms) en situation de connaissance non partagée qu'en situation de connaissance partagée. En outre, à partir de 9 ans, la manière d'introduire les référents varie en fonction de la situation : en connaissance non partagée, les enfants utilisent principalement des formes linguistiques indéfinies alors qu'en situation de connaissance partagée, ils emploient plutôt des formes définies. On constate donc qu'à partir de 9 ans, les enfants sont sensibles à la notion de connaissances partagées : lorsqu'ils savent que leur interlocuteur ne partage pas les mêmes connaissances qu'eux, ils ont conscience de la nécessité d'être plus explicites. Ces travaux indiquent aussi que c'est à 11 ans que les enfants emploient systématiquement les formes indéfinies pour l'introduction des référents, quel que soit le type de situation (connaissance partagée ou non).

Lorsqu'il s'agit du maintien de la référence, l'étude montre que le choix des formes linguistiques utilisées est déterminé par les relations de coréférence : 1) quel que soit l'âge et la situation, les pronoms employés sont principalement coréférentiels, 2) les SN nominaux simples et disloqués sont non coréférentiels. Si la conscience de la coréférence commence à se manifester à l'âge de 9 ans, elle ne devient le principe dominant de l'organisation discursive qu'à partir de 11 ans. Ces recherches appuient le fait que les procédures d'introduction des référents se mettent en place plus précocement que celles qui régissent le maintien de la référence (cf. notamment Hickmann & Schneider, 1993). Les procédures différentes utilisées pour l'introduction ou pour le maintien d'un référent s'expliqueraient par le fait que sur le plan cognitif, l'introduction d'un référent exige moins de coût de traitement. En outre, elle et est plus saillante que le maintien. Sur le plan linguistique, introduire un référent en français met en œuvre des oppositions moins complexes que celles nécessaires au maintien de la référence (Hickmann & *al*., 1995).

Ainsi que nous l'avons vu, la gestion des marques anaphoriques en production dépend non seulement de l'âge mais aussi de divers facteurs tels que le type de texte ou la situation de

connaissances partagées ou non. À ces éléments, il faut ajouter le poids des contraintes exercées par les processus rédactionnels, éléments que nous présentons dans la section suivante.

3.3. La place de la gestion de la cohésion textuelle dans les processus rédactionnels

La production d'un message (écrit ou oral) est une activité complexe et cognitivement coûteuse (Dedeyan & *al.*, 2006). En effet, elle nécessite la mise en œuvre simultanée de différents processus, de l'intention de communiquer à la formulation des mots et phrases prononcés ou écrits, mais également le contrôle du message (Bourdin & Fayol, 2000, 2002). Étudier la production des textes du point de vue du processus rédactionnel est important pour comprendre et analyser les opérations cognitives en jeu dans le cadre de la cohésion et de la cohérence textuelle. Ainsi, comme l'explique Beaudet (2001 : 1), dont les travaux en rédactologie s'intéressent à des textes d'adultes :

> « L'enseignement de la rédaction exige l'adoption du postulat qu'un texte intelligible, compris par son destinataire, possède des qualités, des attributs qu'on peut décrire, en dehors des aptitudes cognitives des lecteurs. Dans cette perspective, la fabrique du sens et sa transmission réussie dans un texte sont redevables d'opérations langagières et discursives ».

D'un point de vue psycholinguistique, Heurley (2006), sur les travaux duquel nous nous appuyons pour cette synthèse, signale que la rédaction d'un texte est généralement conçue comme un processus social, stratégique, dirigé par des buts, soumis à de multiples contraintes, coûteux en ressources cognitives et attentionnelles et qui s'apparente à un processus de résolution de problème décomposable en sous-processus interagissant selon une certaine dynamique. Ainsi, l'acte d'écriture peut être analysé en trois opérations distinctes mais imbriquées les unes dans les autres : la planification, la mise en texte et la révision (Hayes & Flower, 1980). Ces trois composantes peuvent être récursives.

La planification (*planning*) est un processus cognitif qui détermine le but du texte et son plan-guide associé (Fayol, 1984) mais ne se réalise pas forcément par un passage à l'écrit, les opérations pouvant rester mentales. La planification est définie par Hayes & Flower (1981 : 209) de la manière suivante : « Planning is generating content, organizing it, and

setting up goals and procedures for writing [29] ». La planification s'effectue à deux niveaux. D'une part, au niveau de la macro-planification (ou planification rhétorique). Il s'agit de l'élaboration des idées en fonction du but recherché et du public visé. D'autre part, au niveau de la micro-planification qui vise l'organisation structurelle qui conduit à la forme finale du texte. Ainsi, la planification implique toutes les activités cognitives d'écriture nécessaires avant de rédiger un texte. Il est important de souligner ici, à l'instar de Hayes & Flower (1981), que la planification se fait tout au long de l'écriture et que le plan du discours peut ne pas être sous une forme verbale. Il peut se présenter sous la forme d'images mentales ou de schémas écrits.

La mise en texte (*translating*) est l'opération qui consiste à passer des idées à la forme conventionnelle de l'écriture. Hayes & Flower (1981 : 209) la définissent ainsi : « Translating is the act of expressing the content of Planning in written [language][30] ». À ce moment de l'écriture, le scripteur encode sous forme écrite ses représentations mentales. Plus précisément, la mise en texte est la traduction des matériaux activés en mémoire à long terme et organisés par le plan en une mise en forme textuelle acceptable (Fayol, 1984). Les opérations de mise en texte ont pour objet l'organisation d'une suite linéaire articulée de noyaux prédicatifs, en fonction des contraintes imposées par la planification (Fayol & Schneuwly, 1987). Cette opération ajoute de nouvelles contraintes cognitives et force souvent le scripteur à développer, clarifier et/ou réviser son propos. Pour cette raison, il peut exister de nombreux allers-retours entre les opérations de la planification et de la mise en texte.

La révision (*reviewing*)[31] est précisément ce qui se passe lorsque le scripteur évalue son propos et décide d'en changer un élément dans la mise en texte ou la planification. Hayes & Flower (1981 : 209) la définissent ainsi :

> « Reviewing is the act of evaluating either what has been written or what has been planned. When the evaluation of a text or a plan is negative, reviewing often (but not invariably) leads to revision[32] ».

[29] « Planifier c'est générer du contenu, l'organiser et mettre en place des objectifs et des procédures pour écrire » (notre traduction)

[30] « Mettre en texte c'est l'action d'exprimer le contenu de la planification dans le langage écrit » (notre traduction)

[31] Nous attirons l'attention ici sur un fait de traduction : *reviewing* a ici un sens de *retour en arrière*. S'il est traduit ici par *révision*, les auteurs ne mettent pas sur le même plan *reviewing (retour en arrière)* et *revision (révision)*, comme nous pouvons le constater dans leur définition.

Le but de la révision est d'améliorer la qualité du texte (Fayol, 1984) en comparant le texte déjà écrit avec celui que le rédacteur prévoyait d'écrire lors de la planification (Favart & Olive, 2005). La révision peut donc intervenir à tout moment, que ce soit au cours du processus d'écriture, après une erreur par exemple ou à la fin d'une phrase, ou bien à la fin du texte entier. À ce titre, la révision du texte implique d'abord la détection de l'élément à corriger (ou l'évaluation du texte), puis l'identification de sa nature (syntaxique, pragmatique, organisationnelle, etc.), c'est-à-dire l'identification de son (ou ses) inadéquation(s) avec le but poursuivi, et enfin sa correction. La phase de révision est de ce fait très coûteuse en termes cognitifs (Dedeyan & *al.*, 2006).

Notons tout de même ici que cette notion de révision ne recoupe pas les mêmes éléments en fonction des chercheurs. Certains voient dans la révision une modification effective apportée à un texte (Scardamalia & Bereiter, 1983). D'autres chercheurs décrivent la révision comme sous-processus ou composante du processus de l'écriture visant à améliorer le texte déjà écrit et pouvant aboutir ou non à des modifications effectives du texte (Hayes & Flower, 1980, Chesnet & Alamargot, 2005). D'autres chercheurs encore définissent la révision comme une composante du contrôle de la production écrite (Hayes, 1996 ; Roussey & Piolat, 2005), assurant ainsi les fonctions de vérification et d'amélioration du produit fini, la supervision des autres processus (planification des buts, programmation des traitements, etc.) et la suppléance de certains processus défaillants. Pour une revue de ces travaux, nous renvoyons à Heurley (2006).

Les trois processus de l'écriture ainsi décrits ne sont pas un enchaînement normé et devraient plutôt être vus comme des opérations à employer à la demande, en fonction de l'avancement du processus d'écriture (Hayes & Flower, 1981 ; Fayol & Schneuwly, 1987). En effet, si la planification, la mise en texte et la révision sont des processus qui se retrouvent chez tous les scripteurs, ils dépendent toutefois complètement de chacun : certains rédacteurs passent très rapidement à la phase de mise en texte, d'autres préfèrent être certains que leur planification est détaillée au maximum avant de commencer la rédaction.

Les processus rédactionnels exercent sur le système cognitif humain des demandes de traitement considérables (Kellog, 1987) et ne sont donc pas maîtrisés dès les débuts de l'écrit.

---

[32] « Revenir en arrière c'est le fait d'évaluer soit ce qui a été écrit soit ce qui a été planifié. Quand l'évaluation d'un texte ou d'un plan est négative, le retour en arrière mène souvent (mais pas systématiquement) à la révision » (notre traduction)

Favart & Olive (2005) rappellent que lorsque le scripteur commence à rédiger des textes, il met en place une stratégie adaptative qui lui permet de faire face aux demandes de traitement de la rédaction : c'est la stratégie du *knowledge telling* de Bereiter & Scardamalia (1987). Pour ces chercheurs (*ibid.* : 143) :

> « knowledge telling model shows how text generation can take place under these circumstances [given a topic to write about and a familiar genre] without the need for an overall plan or goal, for an elaboration of problem constraints […], or for the problem-solving procedures characteristic of mature composing processes […] »[33]

En d'autres termes, le *knowledge telling* est une planification très locale, par cycles de récupération-production des contenus. Lorsque le scripteur devient expert, il organise les connaissances de manière plus globale, c'est le *knowledge transforming*. C'est seulement grâce à une fluidité croissante des processus au cours de l'acquisition (passage du modèle du *knowledge telling* au *knowledge transforming*) que des ressources peuvent être libérées pour les autres processus (Favart & Olive, 2005), et notamment ceux concernant la gestion de la cohésion textuelle.

Les processus rédactionnels sont donc des opérations complexes qui entrent en jeu lors de la production d'un texte. Le coût qu'ils engendrent dans le traitement cognitif chez l'apprenti scripteur entraîne une diminution des ressources disponibles pour la gestion de la cohérence et de la cohésion textuelle.

Synthèse des notions centrales du chapitre 3

Dans ce chapitre, nous avons voulu montrer que l'acquisition de l'anaphore est non seulement très progressive mais tardive, comme le présentent en particulier les travaux sur la production (Karmiloff-Smith, 1980, 1981, 1983 ; Hickmann 1980, 1984, 1987 ; Bartlett, 1984 ; Bamberg, 1986 ; Wigglesworth, 1990 ; Kail & Hickmann, 1992). Jusqu'à l'âge de 7 ans, voire 9-10 ans en fonction des situations, les enfants éprouvent des difficultés à utiliser correctement les pronoms (personnels et relatifs) et les déterminants (définis et indéfinis).

[33] « Le modèle du *knowledge telling* montre comment la production de texte peut avoir lieu étant donné les circonstances [on donne un sujet sur lequel écrire et un genre connu] sans nécessiter de plan ni d'objectif général, pour élaborer le problème des contraintes […], ou pour les procédures de résolution de problème caractéristiques aux processus de rédaction mature. » (notre traduction)

Dans le maintien de la référence, il semblerait que ce n'est qu'à partir de 6 ans que les enfants commencent à utiliser des pronoms de manière anaphorique. Entre 6 et 8 ans, ils auraient tendance à recourir au pronom pour référer au sujet thématique, en position initiale dans l'énoncé. Après 9-10 ans, le pronom est employé pour référer à d'autres éléments que le sujet thématique. C'est à partir de cet âge également que les enfants sont en mesure d'introduire le référent de manière appropriée et quasi systématique, par le biais d'un SN indéfini.

De surcroît, plusieurs facteurs influencent le choix des expressions référentielles. En fonction du type de texte produit (Bronckart & Schneuwly, 1991 ; De Weck, 1991 ; De Weck & Schneuwly, 1994), les expressions référentielles employées ne seront pas les mêmes : à tous les âges, les SN démonstratifs et indéfinis sont plus présents dans le texte explicatif que narratif, les pronoms personnels anaphoriques sont plus utilisés dans le texte narratif qu'argumentatif… En outre, la connaissance partagée (ou non) par les interlocuteurs a un effet sur le choix des stratégies référentielles (Hickmann, Kail & Roland, 1995 ; Kail & Sanchez y Lopez, 1997) : en connaissance non partagée, les enfants utilisent principalement des formes linguistiques indéfinies alors qu'en situation de connaissance partagée, ils emploient plutôt des formes définies.

Enfin, nous nous sommes intéressés à la question des procédés d'écriture pour tenter de comprendre les enjeux cognitifs à l'œuvre dans ce type de tâche (Hayes & Flower, 1981 ; Fayol & Schneuwly, 1987 ; Heurley, 2006). En effet, la planification, la mise en texte et la révision sont des tâches coûteuses qui s'ajoutent aux contraintes de cohésion textuelle et ensemble, participent à la cohérence globale du texte.

L'acquisition des marques de cohésion étant ainsi progressive et influencée par divers facteurs, dans le chapitre suivant, nous nous interrogeons sur la manière dont l'enseignement les prend en compte.

CHAPITRE 4 : L'enseignement des marques de cohésion et le jugement de cohérence des enseignants

Dans ce chapitre, nous nous intéressons à la dimension didactique des notions de cohérence et de cohésion textuelle. En 4.1, nous abordons la question de la place accordée à l'enseignement de la cohérence et de la cohésion textuelles dans les programmes d'enseignement français. Pour cela, nous cherchons à répondre aux questions suivantes : quand ces notions sont-elles apparues dans les programmes de l'Éducation Nationale ? Comment les préconisations à leurs sujets ont-elles évolué ? Partant des réponses apportées, nous présentons un inventaire des objectifs à atteindre dans l'apprentissage des marques anaphoriques et nous proposons quelques situations didactiques tirées de la recherche dans ce domaine.

Bien que la cohésion textuelle soit au programme de l'école élémentaire en France aujourd'hui, la plupart des enseignants sont démunis et ne savent pas comment l'évaluer. Nous nous penchons donc en 4.2. sur la place de la cohésion dans l'évaluation à travers le jugement de cohérence des enseignants sur les textes d'élèves.

4.1. L'enseignement des marques de cohésion

Après l'étude de la cohésion textuelle des points de vue linguistique et psycholinguistique, nous allons nous intéresser à cette notion sous l'angle didactique, entendue comme : « l'étude des conditions, des modalités de mise en œuvre et des effets d'interventions d'enseignements favorisant les apprentissages dans des classes ordinaires » (Garcia-Debanc, 1990).

La cohérence textuelle est une notion présente dans les programmes français actuels de l'Éducation Nationale (2015). Elle doit être enseignée au cycle 3 (CM1, CM2, 6°), comme le montre l'extrait suivant (Programmes de l'Éducation Nationale, 2015 : 116) :

> « Prendre en compte les normes de l'écrit pour formuler, transcrire et réviser : en lien avec la lecture, prise de conscience des éléments qui assurent la cohérence du texte (connecteurs logiques, temporels, reprises anaphoriques, temps verbaux) pour repérer des dysfonctionnements ».

Pourtant, la cohérence n'a pas toujours été présente dans les instructions officielles et cette notion est une préconisation plutôt récente.

Nous allons voir ainsi en 4.1.1., à travers un tour d'horizon de la place de la notion dans l'histoire des programmes, comment la recherche en linguistique a contribué à faire de la cohérence et de la cohésion textuelles des notions à enseigner. En 4.1.2., nous tentons de recenser quelques propositions d'enseignement émises par les chercheurs pour remédier à ce problème.

4.1.1. La cohérence textuelle dans les programmes d'enseignement

La cohérence textuelle appartient à ce que la tradition scolaire appelle « grammaire ». Depuis les programmes de collège publiés en 1996, la grammaire scolaire distingue trois grands types de grammaire : la grammaire traditionnelle, la grammaire phrastique et la grammaire textuelle. La grammaire traditionnelle et la grammaire phrastique relèvent toutes deux de l'analyse de la phrase mais selon une approche différente. La grammaire textuelle relève de l'analyse du texte. Afin de mieux comprendre quand et comment la cohérence textuelle est apparue dans les programmes d'enseignement, nous proposons de faire un bref historique de l'enseignement de la grammaire scolaire, en nous appuyant principalement sur les travaux de Combettes & Charolles (1999) et Combettes (2016).

4.1.1.1. De la grammaire traditionnelle à la grammaire rénovée du *Plan de Rénovation*

Avant les années 1970, l'enseignement de la grammaire à l'école primaire porte exclusivement sur la phrase dans la tradition de l'analyse logique mise en place à la fin du XIXe siècle (Chervel, 1988, 2006). Cette approche de la grammaire est fondée essentiellement sur la morphosyntaxe et privilégie la décomposition et l'analyse des unités de la phrase, notamment en termes de natures et de fonctions. L'enseignement « traditionnel » de la grammaire, tel qu'il est pratiqué à l'école comme au collège jusque dans les années 1970, est critiqué car jugé trop abstrait et artificiel : l'enseignement de la grammaire ne s'appuie pas sur le fonctionnement réel de la langue mais privilégie les critères sémantiques, entraînant des

confusions entre critères formels et sémantiques, et ne prend pas en compte la hiérarchisation des structures syntaxiques, dans une progression qui va du mot à la phrase (Combettes & Charolles, 1999 ; Combettes, 2016).

Face à ces critiques, le *Plan de rénovation* pour l'enseignement du français à l'école primaire, appelé « Projet Rouchette », est publié en 1971. Ses préconisations prennent appui sur les théories linguistiques issues de l'analyse phrastique, en opposition à celles dont s'inspire la grammaire traditionnelle. Nous faisons remarquer que la distinction entre grammaire « phrastique » et grammaire « traditionnelle » est importante car les deux notions ont souvent été amalgamées par les enseignants. À ce propos, Combettes et Charolles (1999 : 76) indiquent que :

> « Cette assimilation est au demeurant relativement abusive dans la mesure où un certain renouveau grammatical s'était fait jour à partir des concepts empruntés au structuralisme, puis au distributionnalisme, voire au générativisme naissant, renouveau qui concernait il est vrai essentiellement la morphosyntaxe ».

On a donc d'un côté une grammaire « traditionnelle » relevant de l'analyse logique, enseignée avant le *Plan de rénovation* et de l'autre, une grammaire phrastique, « rénovée » (Petitjean, 1999), relevant de l'analyse structurale distributionnelle transformationnelle, préconisée par ce plan. On comprend donc pourquoi le *Plan de rénovation* de 1971 a marqué une rupture en proposant un enseignement rénové de la langue. Bronckart (2016 : 4) en résume ainsi les grandes lignes :

> « ce plan proposait une démarche pédagogique centrée sur des activités créatives visant l'enrichissement des capacités d'expression orales et écrites […], articulée à des activités de réflexion guidée visant la construction de connaissances relatives aux régularités d'organisation de la langue ».

Plus précisément, comme l'explique Combettes (2016 : 2) : « De ces principes découle, de façon logique, un programme qui n'est pas un programme de notions grammaticales, mais bien un programme de langue, avec des exemples d'énoncés illustrant les structures à travailler ». Dans cette perspective, l'enseignement de la grammaire ne met plus l'accent sur les natures et les fonctions, sur la décomposition et l'analyse des unités de la phrase, comme dans l'enseignement traditionnel. Ce sont à présent les formes fondamentales de la phrase (interrogation, négation…) qui doivent être étudiées dans des énoncés, c'est-à-dire l'analyse du fonctionnement linguistique de structures considérées dans leur totalité.

Cependant, les préconisations du plan de rénovation ayant soulevé beaucoup de critiques et de réticences, notamment à cause des désaccords au sujet des théories linguistiques dont il s'inspire, ses préconisations n'ont pas été véritablement prises en compte dans les textes officiels pour l'enseignement. En témoigne le texte officiel de 1972, *Instructions relatives à l'enseignement du français à l'école élémentaire* (1972 : 25) :

> « On s'attachera au rythme, à l'enchaînement des phrases, aux groupes de mots et aux articulations dans la phrase », à travers l'étude des « transformations de phrases (par exemple, d'une affirmative faire une négative ou une interrogative), les substitutions (remplacement d'une structure par une autre, d'un mot ou d'un groupe de mots par un autre sans changement de structure), les déplacements (en particulier de compléments circonstanciels), les adjonctions ou soustractions à la phrase ».

La notion de cohérence textuelle n'est pas abordée dans les programmes. Toutefois, un des objectifs d'apprentissage de la scolarité primaire est l'apprentissage de l'expression orale. En apprenant à parler mieux, les élèves sont plus à même d'écrire mieux car « la tâche de l'élève est de trouver et d'agencer de façon *intelligible* les termes appropriés à sa pensée » (*Instructions,* 1972 : 15, souligné par nous), que ce soit à l'oral ou à l'écrit. De cette manière, « les occasions de rédiger seront fréquemment saisies ; les tentatives faites pour obtenir une formulation orale *satisfaisante* seront souvent reprises par écrit » (*ibid.*). Derrière les termes généraux *intelligible* et *formulation satisfaisante*, le linguiste peut trouver les lointaines prémices de la notion de cohérence textuelle, même si celle-ci reste implicite et ne fait pas l'objet d'un enseignement précis. D'autres éléments qui rappellent la cohérence textuelle sont encore évoqués dans les *Instructions*. Pour faciliter la bonne formation d'une rédaction, les *Instructions* préconisent de proposer aux élèves des activités de « mise en ordre de séries d'images, photographies, documents, questions pour une enquête ou résultats d'enquêtes » afin d'entraîner les élèves à « percevoir des rapports de succession et de liaison » (*ibid.* : 17). Les *rapports de succession et de liaison* sont des éléments constitutifs de la cohérence textuelle, bien que réduits ici aux connecteurs et présentés de manière très succincte, sans explication ou exemple précis.

La priorité donnée à la phrase se retrouve dans la circulaire sur la *Terminologie grammaticale pour l'école élémentaire* de 1976 : la notion est la première à y être mentionnée. Le choix de la terminologie indique une approche traditionnelle de la grammaire dans la tradition de l'analyse logique, avec le terme *formes de la phrase* mais aussi les notions de *classes de mots* et *fonctions* (*ibid.* : 4). Les programmes de 1978 (: 9) s'appuient explicitement sur la terminologie de 1976, dès l'introduction de la rubrique *grammaire* :

> « Les indications ci-après visent à préciser le palier correspondant au cycle élémentaire en ce qui concerne les objectifs impliqués, pour l'ensemble de la scolarité primaire, par la circulaire du 25 octobre 1976 sur la terminologie grammaticale (à laquelle il convient de se reporter) ».

Dans cette circulaire, il n'y a pas de trace de notions liées à la cohérence textuelle. En revanche, dans les programmes de 1978, on retrouve dans la partie *langue écrite* pour le cycle élémentaire (CE1 et CE2) des objectifs ayant trait à la cohérence textuelle, bien que très concis et généraux : à propos d'un récit, l'élève doit « inventer une suite ou des variantes cohérentes » (*ibid.* : 8). De manière générale, tous les textes rédigés par les élèves doivent être appréciés par l'enseignant en prenant en considération la « pertinence et [la] cohérence du contenu par rapport aux intentions du texte produit » (*ibid.* : 9), quel que soit le but du texte (raconter un fait, rédiger un commentaire, résumer une histoire, répondre à des questions, …). Pour le cycle moyen (CM1 et CM2), l'objectif majeur est de « développer chez tous les enfants la capacité de communiquer et de s'exprimer avec aisance, clarté et correction, oralement et par écrit, dans la langue d'aujourd'hui » (*ibid.* : 10). Les activités d'expression écrite doivent donc rendre l'élève capable de « rédiger […] un texte d'une dizaine de lignes répondant à des consignes simples de fond et de forme, […] comprendre la nécessité d'un plan et de commencer à organiser ses rédactions » (*ibid.* : 12). Les textes ainsi produits s'apprécient en prenant en considération l' « adéquation du texte produit, du registre de langue utilisé, aux intentions de l'auteur et à la situation » (*ibid.*). La cohérence est réduite ici à l'organisation des idées dans un texte, sans autre précision.

Dans l'enseignement de la grammaire, aucun changement notable n'intervient non plus dans les *Programmes et instructions à l'école élémentaire* de 1985. La phrase est analysée, « analyse et transformation des formes de la phrase » (*ibid.* : 8) et découpée en unités plus petites, les « éléments constituants de la phrase : le verbe et les groupes fonctionnels » (*ibid.* : 9), tels que le « groupe nominal » et le « groupe verbal ». Mais « rien n'est dit sur l'importance du groupe au niveau de la structuration de la phrase, sur son rôle dans le fonctionnement syntaxique » (Combettes, 2016 : 4). De même, l'analyse grammaticale garde la vision traditionnelle de « distinction des classes de mots », « reconnaissance des fonctions » (Programmes, 1985 : 8). La priorité reste donnée à l'analyse des unités ainsi qu'à leur décomposition, et l'étude des structures et des constructions reste marginale. En ce qui concerne la cohérence textuelle, rien n'est dit, si ce n'est un vague lien avec l'*entraînement à la pratique de la composition écrite* qui regroupe *plan, articulation des idées, organisation en paragraphes* (*ibid.* : 8-9).

4.1.1.2. L'émergence de la grammaire textuelle dans les programmes français de l'école primaire

Dans les années 80 et 90, de nouvelles théories linguistiques commencent à se diffuser dans l'enseignement, amenant une grammaire « reconfigurée » (Petitjean, 1999) qui relève de l'analyse textuelle, en lien avec l'analyse du discours. Comme l'expliquent Combettes & Charolles (1999), l'analyse du discours doit ses débuts à plusieurs théories linguistiques : d'une part, les théories de Benveniste (1974) sur les phénomènes d'énonciation, phénomènes qui distinguent le niveau et le fonctionnement de la phrase de ceux du discours, d'autre part, les théories d'Halliday & Hasan (1976) qui caractérisent les expressions relationnelles entre phrases à même d'assurer au discours une certaine continuité (cohésion et cohérence textuelle). Pour ces linguistes, « le discours ne doit pas être conçu comme une unité grammaticale mais comme « une unité d'une autre espèce : une unité sémantique » » (Combettes et Charolles, 1999 : 82). Partant de cette distinction entre la phrase et le discours ou le texte, a été postulée l'existence de critères d'acceptabilité, critères communément partagés à l'échelle du discours, comme à l'échelle de la phrase. De ce postulat est né le projet d'édifier une grammaire de texte. À l'analyse linguistique du discours, se sont alors ajoutés des critères de l'ordre de l'analyse psycholinguistique, afin de mieux rendre compte des opérations interprétatives en jeu dans la cohérence et la cohésion textuelles. Cependant, un glissement dans l'usage des termes s'est produit dans l'enseignement, comme le font remarquer Combettes & Charolles (1999 : 78) :

> « La [Grammaire Textuelle], dans bon nombre de cas, a été simplement comprise comme une activité grammaticale, plus ou moins classique, « à partir de textes », la nouveauté, s'il y en a une, résidant davantage dans le changement de corpus - de la phrase au texte - que dans une modification quelconque des notions. Étant donnée le vague de l'adjectif « textuel », on comprend que la [grammaire textuelle] ait pu être interprétée comme : « la grammaire à travers les textes » ».

La séparation de l'analyse phrastique et de l'analyse du discours ou du texte a eu pour effet de décrire et nommer différemment les mêmes notions, selon qu'on les envisage du point de vue de l'analyse phrastique ou du point de vue de l'analyse textuelle. Ce point de terminologie constitue un des principaux reproches formulés par les instructions officielles. La circulaire de 1976, *Terminologie grammaticale pour l'école élémentaire,* est éditée afin de transmettre aux enseignants le vocabulaire qu'ils doivent enseigner. Le document justifie son utilité ainsi (*ibid.* : 2) :

> « la simplicité et l'uniformité relatives au vocabulaire grammatical utilisé à l'école élémentaire n'ont pas posé de problème jusqu'au moment où la diffusion de théories linguistiques diverses a amené dans les classes l'emploi de terminologies différentes et parfois concurrentes ».

Les *Programmes et Instructions* de 1985 pour le collège sont les premiers à prendre en compte, de manière prudente, les différents types d'analyse grammaticale (1985 : 26 cité par Combettes & Charolles, 2016 : 5) :

> « [le professeur] montre que les fonctions syntaxiques peuvent être exercées soit par des mots ou groupes de mots, soit par des propositions subordonnées relatives, complétives ou circonstancielles, réunissant ce que l'on a appelé l'analyse grammaticale et l'analyse logique ».

De la même manière, la dimension textuelle et discursive est prise en compte, bien qu'il ne s'agisse pas véritablement d'enseigner l'opposition phrase/texte telle qu'elle était pensée dans les travaux scientifiques mais plutôt de prendre en compte le texte dans son ensemble afin de ne pas faire de la « grammaire pour la grammaire », comme le fait remarquer Combettes (2016 : 5) :

> « Un principe peut guider la progression : en sixième et cinquième, on étudie la phrase simple et la phrase complexe, afin d'assurer des bases fermes, sans exclure l'étude des suites de phrases qui constituent un texte ; en quatrième et en troisième, celle-ci prend la place la plus importante, sans que soient exclus des retours et des compléments à l'étude de la phrase ».

Les programmes pour l'école élémentaire de 1991 restent dans la lignée de ceux de 1985, et préconisent toujours une approche traditionnelle de la grammaire. Toutefois, la notion de cohérence textuelle est présente au cycle 2 (CP, CE1), dans la rubrique intitulée *Langue orale* où on peut lire : « l'enfant doit pouvoir […] rapporter, avec un effort pour introduire de la cohérence dans la relation, un conte, un récit, un évènement vécu, un projet » (*Programmes,* 1991 : 22). En lecture, l'élève doit être capable d' « identifier les personnages d'un récit et les retrouver quels que soient les procédés utilisés pour les désigner (noms, pronoms, surnoms, périphrases…) » (*ibid.*). En d'autres termes, il s'agit de reconnaître les diverses expressions et procédés référentiels employés pour référer à une même occurrence, dans le cadre de la compréhension d'un texte. En revanche, la même opération du côté de la production de texte n'est pas mentionnée.

En 1992, les « Documents d'accompagnement » des programmes, *La Maîtrise de la langue à l'École*, explicitent l'approche textuelle. Si rien n'est dit en ce qui concerne l'analyse structurale distributionnelle de la grammaire de phrase (pas de référence aux constructions,

aux structures de phrase) en revanche, l'approche textuelle est largement commentée. Les notions inhérentes à la cohérence et à la cohésion textuelles notamment (anaphore, progression thématique, etc.), apparaissent dans une rubrique intitulée *Apprendre à organiser un texte*. Au sujet de la cohérence textuelle, la notion apparaît dans la rubrique *Production de texte* comme un objectif à atteindre à la fin du cycle 2 mais elle est présentée de manière générale (*ibid.* : 24) :

> « L'objectif que l'on peut raisonnablement viser pour la fin du cycle II est donc la mise en place de cette double capacité : structurer globalement un texte en l'adaptant à la situation dans laquelle il est produit, maîtriser au fur et à mesure que l'on avance dans le texte, les énoncés produits de façon à ce qu'ils soient compréhensibles par un lecteur, c'est-à-dire un tiers qui, a priori, ne sait rien du contenu du texte et de la situation dans laquelle il a été produit ».

Pour atteindre cet objectif, le document préconise de sérier les apprentissages, en mettant l'accent sur différents aspects du texte et notamment « sur sa structure générale et sur sa cohérence », en étudiant particulièrement « la cohésion de la phrase ou celle des enchaînements » (*ibid.* : 24). L'élève travaille également à une compréhension de plus en plus fine des textes. Il doit pouvoir (*ibid.* : 31) :

> « […] bien interpréter le texte dans sa cohérence propre. C'est cette dernière qui doit être perçue par-delà la suite des mots et des propositions reconnues. Dans un texte, cette cohérence est assurée […] par des pronoms ou des expressions qui sont autant de « reprises » d'éléments déjà énoncés. Bien d'autres réalités linguistiques y participent et doivent donc être progressivement interprétées de manière correcte ».

Notons au passage qu'il s'agit là de la première indication précise des éléments qui assurent la cohérence textuelle : « pronoms ou expressions qui sont autant de « reprises » d'éléments déjà énoncés », bien qu'il reste du vague avec « autres réalités linguistiques » puisque ces réalités linguistiques ne sont pas explicitées.

Au cycle 3, les enseignants sont invités à entrer davantage dans l'observation du système de la langue afin que l'attitude réflexive soit au service des activités de compréhension et de production des textes. Pour la cohérence textuelle, cela se traduit par une étude plus poussée des phénomènes linguistiques et par la distinction entre cohérence et cohésion (*ibid.* : 35) :

> « Il est ainsi possible de distinguer deux grands ensembles de phénomènes qui doivent particulièrement retenir l'attention des maîtres :

- Les faits de langue qui entrent en jeu dans la cohérence et la cohésion du texte (par exemple les anaphores, les connecteurs, les marques d'énonciation, les indices temporels…)
- Les faits de langue qui donnent à la phrase sa grammaticalité (par exemple les phénomènes d'accord, la construction des compléments…) »

Dans le travail de production de texte, une attention particulière est portée à la cohésion textuelle. Le document précise que les marques temporelles, les connecteurs et les anaphores sont des facteurs de cohésion textuelle. Les anaphores sont particulièrement détaillées et doivent faire l'objet d'un apprentissage régulier (*ibid.* : 39) :

> « il conviendra d'être attentif, tout au long du cycle III, au fonctionnement des expressions que les grammairiens appellent des anaphores. […] Les formes les plus courantes de ces reprises par lesquelles se constitue la cohésion du texte sont les pronoms personnels de troisième personne. Les enfants doivent apprendre qu'on ne peut les employer directement sans avoir déjà présenté ce ou celui qu'ils désignent. Ils doivent également apprendre qu'un usage maladroit produit des ambiguïtés que le lecteur ne pourra plus interpréter. ».

Les préconisations de ce document sont largement inspirées des travaux en linguistique de l'époque, qui se sont particulièrement intéressés aux élèves eux-mêmes, tant du point de vue de leur compréhension du discours que de sa production[34]. Pour autant, ce document est plus un outil de formation que de véritables instructions officielles (Petitjean, 1999).

L'analyse textuelle a d'abord été présente dans les documents d'accompagnement de 1992 puis a fait son entrée dans les programmes d'enseignement officiels de 1995 pour l'école élémentaire et 1996 pour le collège, qui marquent ainsi un tournant. En effet, la priorité est donnée à l'analyse du discours et du texte (c'est le premier élément cité) et la coupure est nette entre d'un côté, l'approche textuelle et de l'autre, l'approche phrastique, comme en témoigne l'extrait ci-dessous (*Programmes cycle 2*, 1995 : 27) :

> « Les éléments suivants font l'objet d'un apprentissage progressivement structuré puis systématisé :
>
> - Le texte, le paragraphe, la phrase dans le texte, la ponctuation
> - La phrase simple, la distinction du groupe nominal sujet et du groupe verbal »

De même, les objectifs de fin de cycle 3, qui déterminent les connaissances nécessaires à la maîtrise de la langue, distinguent trois grands domaines de connaissances dans l'enseignement de la grammaire (dans l'ordre) : « l'organisation et la cohérence du texte »,

[34] On peut penser notamment aux travaux de Charolles, Reichler-Béguelin, Schnedecker…

« la phrase », « les classes de mots » (programmes 1995 : 37). Comme l'explique Combettes (2016 : 6) :

> « la distinction souvent reprise par la suite, entre les trois « grammaires », grammaire de discours, grammaire du texte, grammaire de la phrase, sembl[ent] résumer la nouvelle organisation de l'étude de la langue. […] L'approche choisie est en partie le reflet des positions théoriques de l'époque : une coupure nette sépare la syntaxe et le domaine du discours et du texte. Seuls quelques faits de langue ont donc droit à un traitement particulier dans la mesure où ils participent au codage de la cohérence textuelle, les autres étant laissés au domaine de la grammaticalité. C'est ainsi que les types et les formes de phrase sont placés sous la rubrique « discours », en raison des aspects sémantiques et énonciatifs, alors que les classes et les fonctions demeurent du côté de la morphosyntaxe ».

Ce changement d'approche se traduit par l'étude en classe de « l'organisation d'un texte », à l'aide des notions issues de la linguistique textuelle telles que l'opposition récit/discours, les divers types de progressions thématiques ou la distribution des anaphores dans un texte narratif (Masseron, 2016). Au cycle 2, les élèves sont « entraînés à saisir l'organisation générale des phrases et du texte et, pour cela, à repérer notamment les indices typographiques (majuscule, ponctuation), […] et les éléments qui donnent leur cohérence au texte ». (*Programmes,* 1995 : 25). Cependant, aucun détail n'est ajouté concernant ces *éléments qui donnent leur cohérence au texte*.

Dans les programmes de 1995, au cycle 3, l'un des objectifs d'apprentissage est l'approfondissement de la découverte du sens. À cette fin, les exigences en matière de compréhension de texte sont renforcées (*ibid.* : 36), afin :

> « d'accéder à une compréhension fine : bonne connaissance des enchaînements de l'écrit (enchaînements chronologique, logique et analogique, jeu de pronoms, ponctuation qui facilitent l'interprétation du texte), mise en relation des informations prélevées dans le texte, découverte de l'implicite ».

Les enchaînements de l'écrit dont il est question (*chronologique, logique, analogique, jeu de pronoms, ponctuation*) sont caractéristiques de la cohérence et de la cohésion textuelle mais ne sont pas explicitement présentés comme tels. En outre, les termes employés sont peu précis. En revanche, la cohérence du texte apparaît clairement comme une connaissance nécessaire à la maîtrise de la langue orale, de la lecture et de l'expression écrite : « L'organisation et la cohérence du texte : rôle du chapitre, du paragraphe, des mots permettant l'agencement des parties entre elles » (*ibid.* : 37). Mais là encore, la notion de

cohérence reste très peu détaillée. Cependant, comme le précisent Combettes & Charolles (1999 : 79) :

> « (…] il serait injuste d'ignorer les apports de la [grammaire textuelle] dans les parties qui lui sont propres ; des phénomènes comme l'anaphore, la progression thématique, les marques énonciatives, l'opposition des plans, qui n'étaient pas traités dans la grammaire phrastique, ont eu ainsi droit de cité, et la prise en compte de ces points, étroitement liés aux activités de production et de compréhension des textes, peut être considérée comme une avancée ».

Les programmes de 2002 se situent dans le même courant de pensée dominé par l'approche textuelle. Dès le cycle 2, il est recommandé d'articuler activités de lecture et d'écriture car les deux sont complémentaires pour la maîtrise du langage : comprendre des textes (lecture) permet d'acquérir progressivement leur production (écriture). Pour ce qui est de la compréhension de texte, les programmes préconisent que le lecteur « construise des représentations successives de ce qu'il lit et les articule entre elles » (*ibid.* : 47). Cela suppose :

> « que l'on découpe dans le texte des ensembles cohérents d'information et qu'on les mémorise au prix d'un important travail de sélection et de condensation. Cela suppose aussi que l'on traite efficacement toutes les marques qui assurent la cohésion du texte : ponctuation, déterminants, substituts du nom (pronoms, synonymes), connecteurs, marques de temporalité… […]. C'est en attirant [l'attention des élèves], en cours de lecture, sur les marques de cohésion rencontrées et en assurant leur bonne interprétation qu'on leur permet de s'approprier progressivement leur usage ».

La cohésion textuelle est détaillée dans les programmes dès le cycle 2. Au cycle 3, la cohésion ou la cohérence textuelle ne sont pas mentionnées. Toutefois, un des objectifs de l'observation réfléchie de la langue est le repérage d'éléments caractéristiques de la cohésion textuelle, tels que des mots de liaison (connecteurs temporels, spatiaux et logiques) ou les divers substituts d'un nom (pronoms et substituts nominaux) (*ibid.* : 75) dans les textes lus afin que l'élève soit capable de choisir l'élément adéquat lors de la production d'un texte.

En revanche, la notion de cohésion fait explicitement son apparition dans les *Documents d'accompagnement des programmes* de 2002, *Lire et écrire au cycle 3* (2003), qui proposent divers exercices afin de « s'entraîner sur des problèmes de cohésion du texte (marques grammaticales, reprises/substituts, connecteurs » (*ibid.* : 43). De la même manière, ces documents proposent de s'entraîner sur des problèmes de cohérence du texte (chronologie, personnages, temps, lieux, action, énonciation), distinguant ainsi clairement la cohésion et la cohérence textuelle.

4.1.1.3. Le retour de la grammaire phrastique traditionnelle et compromis dans les programmes français publiés en 2008

Dans les programmes de 1995, comme dans ceux de 2002, la mise en avant de « l'Au-delà de la phrase » (Charolles, 1988a) semble faire passer au second plan toute la partie de la morphosyntaxe qui ne peut être rattachée à la problématique du codage de la cohérence textuelle et discursive (Combettes, 2016). Le déséquilibre constaté est un des facteurs qui conduit à de nombreuses critiques de ces textes. En réaction, les programmes d'enseignement de 2008, que ce soit pour l'école élémentaire ou pour le collège, marquent une volte-face avec le retour de l'approche traditionnelle de la grammaire avec son analyse logique (Doquet, 2013). La connaissance et reconnaissance des unités constitutives de la phrase, ainsi que leur analyse en termes de nature et de fonction, est de nouveau mise au premier plan (programmes pour l'école élémentaire 2008 : 14) :

> « L'élève apprend à identifier les types et formes de phrase, les principales classes et fonctions des mots, à maîtriser l'utilisation des temps verbaux, à mémoriser et à appliquer les règles qui prévalent dans la langue française écrite ».

Ou encore pour le collège (programmes pour le collège 2008 :26, cité par Combettes, 2016 : 8) :

> « La reconnaissance des classes de mots (leur *nature*), constitue un préalable indispensable à la maîtrise des accords au sein de la phrase et à l'enrichissement de la syntaxe ».

Le retour à la grammaire traditionnelle marque la priorité donnée par les programmes à une grammaire réflexive, avec des activités de l'ordre de l'analyse, de l'identification. Les programmes ne sont donc pas des programmes de langue, comme le fait remarquer Combettes (2016), car ils ne correspondent pas à un apprentissage progressif des structures linguistiques, caractérisé par des activités d'expression et de compréhension. La grammaire textuelle est elle aussi mise à l'écart : les seules références à des notions relevant de la grammaire de texte et du domaine de la discursivité, notamment à la notion de cohérence textuelle, sont vagues. Au cycle 2 par exemple, il n'est pas question explicitement de cohérence textuelle mais les programmes précisent, dans la rubrique *Écriture*, que les élèves doivent apprendre à rédiger un texte court et que cela passe par « construire et enchaîner des phrases » (*ibid.* : 9). Nous l'avons vu au chapitre 2, la cohérence textuelle est ce qui fait qu'un amas de phrases devient un texte qui a du sens : il s'agit bien de savoir enchaîner des phrases correctement. Les

progressions précisent, sans toutefois donner une définition de la cohérence, qu'au CE1, les élèves doivent pouvoir « concevoir et écrire de manière autonome une phrase simple cohérente, puis plusieurs, puis un texte narratif ou explicatif de 5 à 10 lignes » (*ibid.* : 25).

Au cycle 3 (CE2, CM1, CM2), la notion de cohérence textuelle apparaît dans la rubrique *Étude de Texte* et non plus *Écriture* comme au cycle précédent. L'objectif de l'étude de texte est la compréhension du sens d'un texte par son analyse détaillée (*ibid.* : 13) :

> « [l'analyse précise d'un texte] consiste principalement en l'observation des traits distinctifs qui donnent au texte sa cohérence : titre, organisation en phrases et en paragraphes, rôle de la ponctuation et des mots de liaison, usage des pronoms, temps verbaux, champs lexicaux. »

La cohérence apparaît ici du côté de la réception puisqu'il est question de la compréhension d'un texte.

Dans les progressions proposées au cycle 3, la cohérence n'apparaît plus du côté de la réception mais de celui de la production puisqu'elle est citée dans la rubrique *Rédaction* et non plus dans la rubrique *Étude de texte.* Elle fait l'objet d'un apprentissage progressif (*ibid.* : 29) :

- au CE2, les élèves doivent veiller à la *cohérence temporelle (temps des verbes)* du texte qu'ils produisent. Ils doivent également faire attention à (*ibid.* : 29) :

  > « sa précision (dans la nomination des personnages et par l'usage d'adjectifs qualificatifs), en évitant les répétitions par l'usage de synonymes, et en respectant les contraintes syntaxiques et orthographiques ainsi que la ponctuation »

  En termes linguistiques, les élèves doivent porter une attention particulière à la cohésion textuelle, notamment en ce qui concerne la référence aux personnages.

- au CM1, les élèves doivent pouvoir rédiger des textes courts en veillant à leur *cohérence*, qui regroupe les « pronoms, conjonctions de coordination, mots de liaison, relations temporelles ». Une attention particulière doit être portée à l'évitement des répétitions « par l'emploi des pronoms personnels ». Il s'agit là d'employer des anaphores pronominales et de varier les expressions référentielles.

- au CM2, l'élève doit être capable de (*ibid.* : 29) :

  > « rédiger différents types de textes d'au moins deux paragraphes en veillant à leur cohérence (mots de liaison exprimant les relations temporelles, spatiales ou logiques, pronoms, déterminants, valeur des temps), en évitant les répétitions par l'emploi de divers pronoms et de

> périphrases, et en respectant les contraintes syntaxiques et orthographiques, ainsi que la ponctuation »

Malgré le chemin parcouru, on peut toutefois regretter le fait que la cohérence textuelle, présente dans les programmes de 2008, reste souvent incomplète ou associée à des notions qui relèvent plutôt de la cohésion textuelle, sans véritablement distinguer les deux notions ou expliquer leur lien étroit. Par ailleurs, elle est parfois citée dans les rubriques concernant les compétences rédactionnelles et parfois dans celles concernant les compétences relevant de l'étude de texte. L'instabilité de la cohérence dans les programmes révèle une notion qui a du mal à être identifiée clairement.

Au collège, un des objectifs de fin de cycle est « la rédaction d'un texte narratif correct et cohérent », dont la longueur varie en fonction du niveau de classe (d'une page à plus de deux) (*Programmes d'enseignement du collège*, 2008 : 6, 8, 10, 12). Pour atteindre cet objectif, l'élève doit être entraîné régulièrement à l'écriture et s'habituer « peu à peu à structurer ses écrits en améliorant leur cohérence » (*ibid.* : 6). Pour autant, ce n'est qu'en 4°, dans la rubrique *Étude la langue*, qu'apparaît une « initiation à la grammaire de texte » (*ibid.* : 9) avec l'introduction des notions de connecteurs (spatiaux, temporels, argumentatifs) et en 3°, celles de reprises anaphoriques et de thématisation (*ibid.* : 11).

Le résultat des compromis entre enseignement traditionnel caractérisé par l'analyse logique de la grammaire de phrase et enseignement inspiré de la grammaire de texte est l'ambigüité des objectifs de l'enseignement grammatical. À titre d'exemple, Masseron (2014) reproduit un exercice de manuel scolaire inspiré des programmes de cette période. Il s'agit d'un texte à trous dans lequel l'élève doit combler les blancs pour « construire un texte cohérent ». Mais chaque trou est rempli par une proposition de réponse, comme par exemple « groupe sujet » ou « complément de lieu ». Comme l'explique l'auteure (*Ibid.*), ce type d'exercice illustre l'ambivalence des programmes d'enseignement : la grammaire de phrase (les indications de fonction dans les blancs) est soumise à des règles d'inférence référentielle et de suivi thématique qui, comme le dit la consigne, fondent la cohérence et l'enchaînement des phrases. Comme l'indique Masseron (2014), l'alternance des approches de l'enseignement de la grammaire montre :

> « […] les contradictions d'un modèle d'analyse hétérogène selon qu'il tend à inculquer les règles de l'orthographe ou à sensibiliser les élèves aux formes de reprise et à la cohésion des enchaînements (quel modèle d'analyse retenir pour décrire quels phénomènes ?). »

4.1.1.4. Les programmes d'enseignement de 2015 : une prise en compte équilibrée des « trois grammaires » ?

Les programmes d'enseignement entrés en vigueur à la rentrée 2016 semblent s'écarter de l'approche traditionnelle de la grammaire. Sur le plan méthodologique, que ce soit à l'école élémentaire ou au collège, les activités d'observation de la langue sont préconisées, dans une démarche de découverte qui s'appuie sur l'utilisation de corpus. Ces corpus peuvent être composés de productions d'élèves, dans la mesure où ils permettent de problématiser un fait de langue, comme on le voit dans les deux extraits ci-dessous (Programmes d'enseignement 2015) :

> « [l'étude de la langue] prend appui sur les textes étudiés et sur les textes produits par les élèves ». *(ibid.* : 114)

> « [il est préconisé un] travail à partir des écrits des élèves, acceptables ou non, de corpus de phrases et/ou de textes créés, de textes littéraires ou non pour inviter à la problématisation (situation-problème) ». (*ibid.* : 239)

Sur le plan des contenus des programmes, le changement est également notable. En ce qui concerne la grammaire phrastique, au cycle 4, l'identification des unités de la phrase est mise au second plan pour laisser la priorité à l'étude des structures syntaxiques et à leur fonctionnement. L'objectif ainsi visé est de découvrir et étudier les propriétés particulières de telle ou telle structure.

Pour la grammaire textuelle, le cycle 2, en qualité de cycle des apprentissages fondamentaux, donne la priorité à l'apprentissage de la langue française. Aussi, l'apprentissage de la lecture et de l'écriture reste primordial. La première mention de la cohérence apparaît dans les attendus de fin de cycle en écriture (*ibid.* : 20) : « rédiger un texte d'environ une demi-page, cohérent, organisé, ponctué, pertinent par rapport à la visée et au destinataire ». À cet attendu, présenté de manière générale, s'associent plusieurs connaissances et compétences qui font appel à la notion de cohérence : le premier jet de l'écriture et la révision du texte produit[35].

Lors du premier jet d'écriture, les programmes préconisent la « mise en œuvre (guidée puis autonome) d'une démarche de production de textes : trouver et organiser des idées,

[35] Notons au passage l'emploi du terme *révision* qui renvoie directement au modèle du processus rédactionnel de Hayes & Flower (cf. ce chapitre, section 3.1.4.).

élaborer des phrases qui s'enchaînent avec cohérence, écrire ces phrases » (*ibid.* : 21). Pour cela, le programme suggère de passer par la « recherche collective des caractéristiques attendues du texte à produire » (*ibid.* : 21). En ce qui concerne la révision des écrits produits, il est attendu des élèves qu'ils repèrent les « […] dysfonctionnements dans les textes produits (omissions, incohérences, redites...) » (*ibid.* : 22) afin de s'appuyer dessus pour améliorer leurs textes. Dans cet objectif, les programmes proposent à l'enseignant de mener des « activités permettant de s'habituer à intervenir sur des textes, notamment modifications de textes (enrichissement de phrases, recours aux pronoms pour éviter les répétitions, changement ou ajout d'éléments...) d'abord effectuées de façon collective [...] puis individuelle [...] » *(ibid. : 22).* Les modifications proposées, notamment le recours aux pronoms pour éviter les répétitions ou le changement/l'ajout d'éléments relèvent de la notion de cohérence et de cohésion textuelle, même si ce n'est pas explicitement indiqué.

L'accès à la langue écrite, que ce soit du point de vue de la réception ou de la production, s'accompagne de l'étude du fonctionnement de la langue. Les programmes d'enseignement rapprochent l'étude du fonctionnement de la langue de la grammaire, de l'orthographe et du lexique, dans une perspective d'enseignement de la grammaire de phrase. Dans ce contexte, la seule mention de la cohérence est réduite à la cohérence sémantique, qui passe par l'identification des « principaux constituants d'une phrase simple en relation avec sa cohérence sémantique (de quoi on parle, ce qu'on en dit) » (*Ibid.* : 24). Il s'agit alors pour l'élève de savoir identifier le groupe nominal, le verbe, les classes de mots, les phrases affirmatives et négatives, la ponctuation de fin de phrases et les signes du discours rapporté. Ainsi au cycle 2, la notion de cohérence textuelle est présente en tant que composante de l'écrit et fait l'objet d'un apprentissage non explicite et plutôt intuitif à travers des activités collectives de réflexion sur les règles de bonne formation des textes.

Le cycle 3 est le cycle de consolidation et à ce titre, il a pour objectif d'affermir et stabiliser les apprentissages ultérieurs. Comme dans le cycle précédent, la notion de cohérence textuelle apparaît comme une compétence de l'ordre de l'activité d'écriture puisqu'un des objectifs de ce cycle (*ibid.* 112) est :

> « la mise en œuvre (guidée puis autonome) d'une démarche de production de textes : convoquer un univers de référence, un matériau linguistique, trouver et organiser des idées, élaborer des phrases, les enchaîner avec cohérence, élaborer des paragraphes ou d'autres normes d'organisation textuelles ».

Le traitement est le même que celui préconisé dans le cycle précédent : la cohérence textuelle ne fait pas l'objet d'un temps d'apprentissage spécifique mais d'une « réflexion préparatoire, collective ou en sous-groupe, sur l'écrit attendu et sur les différentes stratégies d'écriture » (*ibid.* : 112).

Toutefois, au cycle 3, le terme de cohérence s'accompagne d'éléments d'explicitation : les élèves doivent « en lien avec la lecture, [prendre] conscience des éléments qui assurent la cohérence du texte (connecteurs logiques, temporels, reprises anaphoriques, temps verbaux) pour repérer des dysfonctionnements » (*ibid.* 113). Les termes utilisés, *connecteurs, reprises anaphoriques,* sont directement empruntés à la linguistique. Ils sont le signe d'une activité métalinguistique qui s'exerce à travers une étude explicite et réflexive des textes. De même que pour le cycle 2, la rubrique *Étude de la langue* fait uniquement mention de la cohérence sémantique, qu'elle rapproche de la grammaire.

Le cycle 4, cycle des approfondissements regroupant les trois dernières années du collège, distingue deux volets dans l'enseignement du français : d'un côté les compétences langagières et linguistiques et de l'autre la constitution d'une culture littéraire. Encore une fois, la notion de cohérence apparaît dans les attendus de fin de cycle concernant l'écriture, comme l'indique l'extrait suivant (*ibid.* : 232) :

> « en réponse à une consigne d'écriture, produire un écrit d'invention s'inscrivant dans un genre littéraire du programme, en s'assurant de sa cohérence et en respectant les principales normes de la langue écrite »

Cependant, si on veut des précisions, il faut se reporter à la rubrique *Compétences linguistiques : étude de la langue (grammaire, orthographe, lexique)*[36] (*ibid.* : 236). On retrouve donc la notion de cohérence textuelle dans l'étude de la langue ainsi que la notion de cohésion (*ibid* : 236) :

> « La grammaire [est] au service des compétences langagières de lecture et d'écriture nécessaires pour s'approprier le sens des textes et mener des analyses littéraires étayées. Ces compétences interviennent également dans la production d'écrits. Les notions concernant la cohérence et la cohésion des textes sont étudiées en contexte, lors des activités de lecture et d'écriture. La progression prend appui sur les écrits des élèves. »

[36] Notons au passage l'ajout du terme *Compétences linguistiques* à une rubrique qui jusque-là s'intitulait uniquement *Étude de la langue (grammaire, orthographe, lexique).*

La notion de cohésion textuelle n'apparaît donc explicitement qu'au cycle 4, alors qu'au cycle 3, les termes linguistiques employés relevaient déjà de la cohésion textuelle, sans toutefois que le terme *cohésion textuelle* apparaissent. En outre, c'est la première fois dans les programmes 2015 que la distinction entre les termes de *cohérence* et de *cohésion* apparaît.

Par *cohésion textuelle*, les programmes entendent des éléments linguistiques tels que « substituts nominaux et pronoms de reprise ; procédés de désignation et de caractérisation, rôle des déterminants indéfinis et définis ; indicateurs de temps et de lieu, de relations logiques ; système des temps » (*ibid.* : 242). Lorsque l'on confronte les termes employés ci-dessus à ceux utilisés pour le cycle 3, à savoir *connecteurs, reprises anaphoriques*, on constate une moindre précision alors même que la notion linguistique à laquelle ils appartiennent est plus précise au cycle 4 qu'au cycle 3.

Lorsqu'il s'agit de la cohérence textuelle, les programmes formulent des éléments plus précis : « maîtrise de la chaîne anaphorique et des substituts nominaux et pronominaux, maîtrise des temps et modes verbaux » (*ibid*. 244). On remarque que le terme de *substituts nominaux* est utilisé à la fois pour la cohésion textuelle et la cohérence. D'autre part, les chaînes anaphoriques renvoient à la référence et de ce fait, à la cohésion textuelle et non à la cohérence.

Au cycle 4 apparaît également une nouvelle notion : celle de progression thématique. L'observation de la progression thématique d'un texte est un objectif d'apprentissage : « représentation schématique de la progression du texte (thème-propos) ; écriture à partir d'une forme de progression imposée » (*ibid.* : 243). En relation directe avec la notion de cohérence textuelle, la progression thématique n'est pourtant pas identifiée comme telle dans les programmes mais comme un élément à part entière, sans lien direct avec la notion de cohérence (*ibid.* : 244) :

> « Les notions à travailler au niveau du texte dans les productions d'écrit des élèves sont les suivantes : la cohérence textuelle (maîtrise de la chaine anaphorique et des substituts nominaux et pronominaux, maîtrise des temps et modes verbaux), l'enchaînement interphrastique (liens logiques), la maîtrise du thème et du propos avec un usage pertinent de la ponctuation. Ces notions sont abordées par approfondissements successifs tout au long du cycle, en s'appuyant sur les réalisations langagières des élèves ».

Les programmes de 2015 ont le mérite de faire apparaître la notion de cohérence textuelle dès le cycle 2, même si elle n'est pas enseignée explicitement comme telle. La

notion de cohésion, elle, apparaît au cycle 4. Les termes employés dans les programmes pour définir les notions de cohérence et de cohésion montrent des similitudes (terme *substituts* employé pour l'une et l'autre) mais aussi des contradictions (éléments linguistiques caractéristiques de l'une tel que *reprise anaphorique* présent dans la définition de l'autre). En outre, la place de ces notions dans les programmes varie : tantôt elles sont considérées comme des compétences relevant de l'activité d'écriture, correspondant au caractère implicite de cette approche aux cycles 2 et 3, tantôt comme des compétences relevant de l'étude de la langue, correspondant au caractère métalinguistique de l'étude de ces phénomènes au cycle 4.

Au terme de ce bref historique, nous pouvons dire que la notion de cohérence textuelle est apparue de manière explicite dans les programmes à partir du moment où la grammaire textuelle a pris le pas sur les grammaires phrastique et traditionnelle. Avant cela, le terme de *cohérence* était employé (cf. programmes de 1978, 1985) mais réduit à l'organisation des idées dans un texte, sans autre précision. Il faut attendre les documents d'accompagnement des programmes de 1992 pour que l'approche textuelle soit clarifiée et surtout, que les notions de cohérence et de cohésion textuelles soient commentées et explicitées. Les programmes d'enseignement des années suivantes (1995, 2002) donnent la priorité à l'analyse textuelle et aux notions de cohérence et de cohésion, bien que n'étant pas autant explicitées que dans les documents d'accompagnement. Toutefois, les programmes de 2008 marquent un retour en arrière dans la mesure où ils font passer l'analyse textuelle et de ce fait, les notions de cohérence et de cohésion, au second plan. En revanche, les programmes de 2015 montrent une volonté d'équilibre entre les différentes approches de la grammaire. Les notions de cohérence et de cohésion textuelles y sont définies en lien l'une avec l'autre et constituent des compétences à travailler dans chacun des cycles.

4.1.1.5. La place de la cohérence/cohésion dans les programmes d'enseignement de la grammaire de quelques pays francophones

Québec

Au Québec, l'enseignement du français a également connu plusieurs phases, comme l'indique la synthèse de Chartrand (2011) sur laquelle nous nous appuyons. Jusque dans les années 1970, les programmes scolaires, appelés *programmes-cadres*, se résumaient à un balisage général. Les premiers programmes de français ont été publiés en 1980, rompant avec

la tradition d'un enseignement jugé trop formaliste et éloigné des pratiques langagières. Notons au passage que ce constat est le même que celui des années 70 en France, qui a conduit au Plan de rénovation de l'enseignement du français (Projet Rouchette). Avec cette nouvelle vision de la langue au Québec, la priorité a été donnée à la langue comme moyen de communication et la grammaire renvoyée au second plan, utile à seule fin de résoudre une difficulté dans la communication écrite. Cependant, ces programmes, malgré une volonté de rupture, n'ont eu que peu d'effet dans les classes. Du point de vue de l'étude de la langue, ces programmes ont eu le mérite d'amener les élèves à appréhender les textes dans leur contexte de production et de réception. En revanche, l'approche communicative préconisée a eu pour effet d'évacuer l'étude de leur textualité (Chartrand, 2011 : 48).

Suite à ce constat, les programmes québécois de 1995 adoptent une approche fondée sur l'analyse des séquences textuelles. Avec cette rénovation de l'enseignement de la grammaire, un nouveau regard sur la langue et sur l'analyse du français a été proposé, à travers une observation de la langue qui prend en compte à la fois la grammaire de phrase et la grammaire de texte, comme l'indiquent les programmes : « observer le fonctionnement d'un phénomène grammatical afin de découvrir les règles d'organisation et de fonctionnement dans les textes lus » (Chartrand & Paret, 2010 : 63), et « évaluer l'apport du phénomène grammatical étudié au sens de la phrase et du texte dans les textes lus et rédigés par les élèves » (*ibid.* : 63). Ces programmes s'inscrivent dans un courant d'enseignement de la grammaire qui existe depuis plus de 40 ans en Europe.

Une nouvelle réforme est venue réorganiser les programmes d'enseignement en 2005-2007, en formulant les contenus de formation en termes de *compétences*. Cependant, en ce qui concerne l'enseignement de la grammaire (ou *travail sur la langue*), les contenus restent les mêmes que dans les programmes précédents. Ainsi, au sujet de l'enseignement du français au Québec, Chartrand (2011 : 50) résume la situation ainsi :

> « En somme, pour l'enseignement grammatical, les enseignants de français ont eu trois programmes en 25 ans. […] Celui de 1980 prescrivait de mettre la grammaire au service de la communication, sans préciser comment (chaque enseignant devait le découvrir !), préconisait une grammaire à l'occasion, n'imposait pas de contenus précis, ne présentait pas explicitement une critique des contenus et démarches de l'enseignement traditionnel de la grammaire, pourtant en bonne partie incompatible avec ses orientations. Celui de 1995 prônait un *enseignement systématique* de la grammaire (en continuité avec les années 1950 et 1960 et en rupture avec celui de 1980) où les contenus étaient organisés selon la logique du système de la langue, plutôt que celle des besoins langagiers des élèves. Celui de 2005-2007, mise à part la référence au programme antérieur, ne s'engage

pas. Il se contente de dresser une liste de « notions et concepts » grammaticaux à enseigner sans préciser en quoi ils sont utiles ni quand les enseigner ».

Suisse romande

En Suisse romande, une rénovation de l'enseignement a également eu lieu dans les années 70-80 et a touché les contenus et les méthodes d'enseignement du français à l'école obligatoire (Reichler-Béguelin, 1999). De même que le *Plan de rénovation* français, dont il s'inspire, le document *Maîtrise du français* (1979) marque une rupture dans l'enseignement suisse en donnant la priorité à une visée communicative de la langue, à travers quatre grandes finalités : savoir écouter, savoir parler, savoir lire et savoir écrire (Simard & *al.*, 2010). Comme l'explique Bronckart (2016), de ces préconisations découle un ensemble de travaux visant à la modernisation de l'enseignement de la langue maternelle :

- l'élaboration d'une méthodologie générale intégrant et coordonnant les traditionnelles composantes d'expression, grammaire, vocabulaire et conjugaison ;
- la production de grammaires et d'autres moyens d'enseignement adaptés aux différents niveaux de la scolarité ;
- la mise en place d'un dispositif de suivi et d'évaluation des effets de l'implémentation des innovations, assurés par des formateurs, linguistes et pédagogues réunis dans le groupe Bally.

À travers les préconisations de ce document, l'enseignement de la grammaire visait à écarter toute caractéristique de la grammaire traditionnelle, en abandonnant par exemple des définitions « qui mélang[ent] critères référentiels et critères morphosyntaxiques » ou en rejetant « une didactique procédant par mémorisation de définitions et de règles, suivie de leur application » (Bulea Bronckart & Bronckart, 2017 : 23). L'enseignement rénové du français qui est proposé alors prend notamment appui sur l'analyse distributionnelle et la grammaire générative transformationnelle, comme l'indique Bronckart (2016 : 4) :

> « Ces tentatives se sont caractérisées aussi, sur le plan de l'identification et de la dénomination des entités linguistiques, par l'instauration d'une démarche d'abord fondée sur des procédés (manipulations) et critères syntaxiques ».

Dans les années 90, la grammaire de l'oral et la grammaire textuelle font leur apparition dans les textes officiels avec les notions de *connecteurs*, *anaphores*, *genres* et de nouvelles procédures de travail (résolution de problèmes d'écriture, lecture suivie, ateliers de

négociation graphique, etc.) (Aeby, De Pietro & Wirthner, 2001). Mais la grammaire textuelle et la grammaire de phrase restent des domaines étanches et superposés (Reichler-Béguelin, 1999), bien qu'enseignées toutes deux. Comme l'indiquent Buléa Bronckart & Bronckart (2017), la *Conférence Intercantonale de l'Instruction publique de la Suisse romande et du Tessin (*CIIP) a engagé dans les années 2000 un réexamen d'ensemble des programmes et méthodes d'enseignement du français à l'école obligatoire. Un travail de fond a été entrepris qui a donné lieu à la publication d'*Orientations* (CIIP, 2006) énonçant les finalités et les principes généraux de la discipline, puis en l'adoption et la diffusion d'un *Plan d'Etudes romand* (CIIP, 2010) qui inventorie les objectifs d'apprentissage et les contenus d'enseignement. Les orientations qui découlent de ce remaniement restent dans le prolongement des préconisations du document fondateur de 1979, *Maîtrise du français,* mais proposent une actualisation des fondements de l'enseignement du français. Cette actualisation se traduit notamment par l'enseignement de la production et la compréhension de textes, qui doit être le « point de départ pour construire et apprendre les concepts grammaticaux » et le « point d'arrivée pour les mettre en œuvre dans la lecture et l'écriture » (*Orientations* : 11, cité par Buléa Bronckart & Bronckart, 2017). Ainsi, l'enseignement de la grammaire en Suisse romande se veut large, incluant à la fois l'approche phrastique dans son analyse logique et structurale (types et formes de phrases, classes grammaticales, fonctions syntaxiques…) et l'approche textuelle.

L'enseignement de la grammaire du français, que ce soit en France ou dans d'autres pays de la francophonie, a ainsi connu plusieurs phases. Les différentes théories linguistiques qui l'ont inspiré semblent se retrouver dans les programmes actuels, notamment en France et en Suisse, sans s'exclure les unes les autres, comme cela a longtemps été le cas. L'enseignement explicite de la cohérence et la cohésion textuelle est relativement récent mais reste souvent peu détaillé et exprimé en des termes linguistiques parfois opaques pour les enseignants, voire confus. La complexité de la définition des notions de cohérence et cohésion, même pour les linguistes, n'est sans doute pas étrangère à ce phénomène. En outre, nombres d'adultes aujourd'hui enseignants ont vécu leur scolarité dans la prédominance de l'enseignement traditionnel de la grammaire. De ce fait, il n'est pas étonnant que beaucoup de professeurs se trouvent démunis lorsque vient le temps d'enseigner à leurs élèves à « rédiger un texte cohérent ».

4.1.2. L'enseignement des marqueurs référentiels

Lorsqu'il est question de cohérence ou de cohésion textuelle dans les programmes, et plus particulièrement lorsqu'il s'agit de réinstancier un référent dans un texte d'élève, beaucoup d'enseignants réduisent cette notion à l'évitement des répétitions (Schnedecker, 1995). Or, comme le résume Garcia-Debanc (1993 : 23), l'enseignement des marqueurs référentiels nécessite « d'inventer des démarches didactiques s'appuyant au plus près sur les besoins constatés chez les élèves en production d'écrits et fondées sur une analyse linguistique précise des obstacles ». L'enseignement des marqueurs référentiels peut donc s'effectuer en plusieurs étapes.

En premier lieu, il s'agit d'identifier les problèmes rencontrés par les élèves lors de la production d'écrit, et plus précisément dans le cadre de notre recherche, les difficultés ayant trait à la référence. Plusieurs chercheurs ont travaillé avec cet objectif, couplé ou non à des propositions didactiques[37].

Une fois les difficultés identifiées, il est plus facile de mettre en œuvre une démarche didactique efficace pour l'enseignement des notions qui posent problème (Tauveron, 1987 ; Djebbour & Lartigue, 1994). Ainsi, quelle que soit la forme de travail envisagée pour améliorer les compétences discursives des élèves en matière d'anaphores, les objectifs d'apprentissage restent les mêmes. En effet, après avoir mis en place progressivement la terminologie métalinguistique adéquate pour parler du texte et s'y référer, il est nécessaire de dépasser le traitement local des faits linguistiques et de prendre en compte leur dimension discursive. En d'autres termes, les élèves doivent être sensibilisés au fait qu'un remaniement ponctuel (supprimer une répétition et la remplacer par un pronom par exemple) peut avoir des répercussions sur la cohérence du texte entier (Djebbour & Lartigue, 1994).

Un autre axe de travail concerne les remaniements syntaxiques (*Ibid.*). L'élève doit apprendre que les problèmes de répétitions conduisent à des transformations syntaxiques (effacement du sujet par exemple). L'évitement des répétitions peut également conduire à une condensation de deux phrases en une ou à un déplacement de l'information, opérations qu'il convient de prendre en compte afin d'une part, que le lecteur comprenne toujours de quoi il

[37] On peut penser notamment à Tauveron, 1987 ; Charolles, 1988a, 1988c ; Reichler-Béguelin, 1988b, 1995 ; Garcia-Debanc, 1993 ; Djebbour & Lartigue, 1994 ; Schnedecker, 1995 ; Cappeau & Roubaud, 2005 ; Cavanagh, 2006.

s'agit et d'autre part, que la lecture soit rendue plus fluide, le texte plus « cohérent ». Enfin, il s'agit également de faire prendre conscience aux élèves de la diversité des substituts et de la possibilité d'en utiliser plusieurs dans un texte (Tauveron, 1987 ; Djebbour & Lartigue, 1994 ; Paret, 2003). Garcia-Debanc (1993) propose un inventaire des objectifs à maîtriser dans le domaine de la référence et une carte des activités correspondantes, en s'appuyant sur Reichler-Béguelin (1993). Tous les objectifs d'apprentissage précédemment cités peuvent être travaillés en prenant pour supports des textes qui en favorisent l'acquisition progressive, mettant ainsi le texte au centre de l'apprentissage de la langue (Chartrand, 2008 ; Marmy-Cusin, 2012), dans une tentative de grammaire intégrée qui associe à la fois la grammaire phrastique et la grammaire textuelle (cf. section 4.1.1).

Dans la même période, au milieu des années 1990, Djebbour & Lartigue (1994) ont travaillé sur des activités de construction de savoirs métalinguistiques dans des classes de cycle 3 (CM1 et CM2) pour améliorer les compétences discursives des élèves en matière de référence. Partant de problèmes rencontrés par les élèves au niveau de la désignation des personnages, c'est-à-dire au niveau des anaphores, dans des activités de lecture et d'écriture de textes narratifs, elles cherchent à mettre en place des projets d'enseignements pour résoudre ces problèmes. À cet effet, les chercheuses proposent trois situations didactiques que nous présentons ci-après, à la fois dans le but de dresser un état des lieux des savoirs métalinguistiques des élèves, et de proposer un projet d'enseignement de ces notions.

La première situation didactique que ces chercheuses préconisent est le repérage des chaînes anaphoriques (voir aussi Lentz, 1993 ; Bucheton 1994). Dans un premier temps, ce repérage est effectué dans des textes d'auteurs. L'objectif de cette situation est de faire réfléchir les élèves sur les personnages : leur introduction dans le récit, la manière dont ils sont situés les uns par rapport aux autres. Pour aider à ce repérage, les élèves doivent surligner d'une même couleur chacun des éléments d'une même chaîne. Dans un second temps, ce repérage par le surlignage en couleur intervient dans les récits produits par les élèves eux-mêmes. L'objectif poursuivi ici est de permettre la mise à distance du texte pour une prise de conscience des éventuels problèmes anaphoriques (ambiguïté, répétitions abusives…). On retrouve cette première phase de reconnaissance dans une situation décrite par Tauveron (1987), du moins dans son objectif car la mise en œuvre est différente du fait de l'âge des élèves. Comme l'indique Tauveron (*ibid.*), dans le but de faire des reprises anaphoriques un objectif d'apprentissage, trois enseignantes de CP proposent à leurs élèves de souligner dans un texte tous les pronoms personnels et déterminants possessifs. Après leur avoir demandé ce

que cette opération permet de comprendre du texte, les élèves doivent reprendre chaque terme pour remonter à sa source. Autrement dit, il s'agit pour eux de comprendre à qui ou à quoi chaque mot souligné fait référence. Une fois le référent des pronoms retrouvé, les élèves identifient tous les autres termes référant à la même entité (noms propres, substituts lexicaux, etc.), sous la forme d'un jeu de lecture dont la consigne est la suivante : « je peux venir montrer que j'ai reconnu un « il » ou « elle », un « son »… à la seule condition de dire en même temps ce qui se cache derrière ». À l'aide de couleurs différentes, est ainsi dégagée la chaîne de référence de chacun des personnages. Cette première situation, quelle que soit sa mise en œuvre, permet aux enfants, en fonction de leur âge, d'identifier, différencier et classer les moyens linguistiques permettant la reprise anaphorique. Ces moyens sont une aide à la production d'écrit mais également à la lecture car ils permettent aux enfants de mieux comprendre la progression des expressions référentielles dans un texte.

La deuxième situation proposée par Djebbour & Lartigue (1994), (mentionnée aussi chez Tauveron, 1987 ; Charolles, 1988c ; Bergeron, 2003) est la réécriture d'un texte de pair ou d'un texte d'auteur présentant des ambiguïtés référentielles. Sur la base d'un texte d'élève produit dans la classe et présentant de nombreuses répétitions, une évaluation collective est proposée. Par groupes, les élèves font le même travail de repérage que dans la première situation puis proposent une réécriture du texte dans le but de remédier aux dysfonctionnements observés.

La troisième situation, sur la base des conclusions qui ont émergé des deux premières situations, consiste en la rédaction par les élèves d'une fiche-outil. Cette fiche doit contenir des propositions linguistiques pour résoudre des problèmes d'ordre anaphorique, ainsi que les précautions à prendre pour opérer les remaniements dans un texte.

Avec le même objectif d'étude des problèmes référentiels rencontrés par les élèves lors de la production d'un texte, Schnedecker (1995) analyse divers textes d'élèves du CM à la troisième. Pour chaque type de problème référentiel rencontré, elle s'attache à proposer une intervention didactique. Par exemple, lorsqu'il s'agit de renforcer un savoir-faire établi, tel que la désignation d'un référent saillant, ou de sensibiliser aux problèmes que suscite un texte pluriréférentiel, un remaniement du texte peut être envisagé. Le texte d'origine peut être modifié de telle sorte qu'il comporte alors des blancs après un pronom renvoyant à l'un des référents. Pour identifier le référent du pronom, l'élève doit se demander à qui (à quel personnage/quelle entité) renvoie le pronom en question. Un deuxième type d'exercice peut

consister dans le fait de proposer aux élèves un corpus de textes simples (par exemple : résumés de films, extraits de romans…) puis de les sensibiliser aux critères qui aident à déterminer quel est le personnage central, comme l'ordre des mentions dans le texte ou le nombre de reprises. À partir de cette activité, un travail plus précis sur les ambiguïtés et leurs différents degrés (Charolles, 1988a) peut être débuté, tout d'abord à travers une série d'énoncés décontextualisés qui présentent des ambiguïtés potentielles. Dans un second temps, les énoncés réinsérés dans leur contexte d'origine font l'objet d'une discussion autour des éléments qui empêchent justement cette ambiguïté.

Lorsqu'il s'agit d'attirer l'attention des élèves sur le choix des expressions référentielles dans le contexte de la redénomination, il peut être proposé aux élèves une version « pronominalisée » d'un texte. Ils devront alors dire s'ils jugent le texte cohérent, comment ils interprètent la redondance des pronoms, et par quels autres éléments ils pourraient les remplacer (voir aussi Charolles, 1988c). L'accent peut être mis également sur le fait que deux expressions référentielles identiques qui apparaitraient dans deux propositions qui se suivent ne sont pas incompatibles, à la condition qu'elles se manifestent dans deux espaces de véridiction bien distincts et circonscrits. À cet effet, les élèves peuvent être confrontés à un exercice consistant à lever une contradiction sans remplacer l'expression référentielle : *Pierre est un excellent cuisinier. Pierre est un cuisinier incompétent.* Devient : *Dans son restaurant, Pierre est un excellent cuisinier. Chez lui, Pierre est un cuisinier incompétent*. Ou encore : *Marie prétend que Pierre est un excellent cuisinier. En réalité, Pierre est un cuisinier incompétent.*

Ces diverses situations didactiques peuvent être considérées comme des propositions de séances qui pourraient être menées en classe afin d'améliorer les compétences discursives des élèves en matière de référence.

4.2. Évaluer la cohérence et la cohésion textuelles : la place de la cohésion et de la cohérence dans l'évaluation et le jugement de cohérence des enseignants

La cohérence et la cohésion textuelles, si elles font l'objet d'un apprentissage, doivent pouvoir également être évaluées. En 4.2.1, nous tentons de définir la place de la cohésion et

de la cohérence textuelles dans l'évaluation des textes des élèves et nous proposons quelques critères d'analyse issus de la recherche. Chemin faisant, nous constatons que la cohérence textuelle est aussi difficile à évaluer qu'à définir car elle fait appel au jugement de cohérence des enseignants. Or, comme l'indique Rondelli (2010b : 63) : « mettre en œuvre une compétence professionnelle permettant de juger la cohérence d'un texte veut dire posséder, implicitement ou explicitement, une définition de la textualité ». En 4.2.2., nous montrons que les enseignants ne maîtrisent qu'imparfaitement cette notion et posent sur les textes de leurs élèves un jugement de cohérence incomplet.

4.2.1. L'évaluation des productions écrites d'élèves

Avant d'aborder la question de l'évaluation de la cohérence et de la cohésion textuelles, il semble pertinent de faire un point général sur l'évaluation des productions écrites des élèves. Charolles (1978) décrit ainsi le mécanisme de l'évaluation, tant du point de vue phrastique que textuel : pour qu'une suite de mots soit admise comme phrase par un locuteur, il faut qu'elle respecte l'ordre institué par le système de la langue dans laquelle la phrase est exprimée. Tout membre d'une communauté linguistique « a une connaissance intuitive et une pratique immédiate de ces contraintes structurelles » (Charolles, 1978 : 7). Il en résulte une *norme minimale* (*ibid.*) théorisée par la grammaire, à partir de laquelle tout locuteur, et *a fortiori* les enseignants, peut distinguer les éléments corrects et les éléments inappropriés, ce « qui est français » et ce qui ne l'est pas.

Pour étudier la manière générale dont les enseignants évaluent les productions écrites des élèves, Halté (1989) propose à un échantillon d'enseignants du second degré (une centaine) d'annoter un texte narratif d'élève de 5°, « la copie de Cécile »[38]. Les annotations, traces de l'évaluation du texte par les enseignants, sont ensuite analysées par le chercheur. Le premier constat concerne les écarts à la norme, plus souvent visés, au détriment des compétences mettant en jeu l'invention : « les annotations portent massivement plus sur « la forme » que sur « le fond ». » (Halté, 1989 : 17). Ce constat était déjà présent dans les travaux de Charolles (1978) qui s'intéressaient aux « interventions des maîtres sur les énoncés écrits malformés des élèves » (Charolles, 1978 : 10). Du point de vue de la cohérence textuelle, la

[38] Texte narratif issu d'une expérience ultérieure de l'auteur : voir « Ecrire un roman historique en classe de 5°. », Halté, J-F., 1982, *Pratiques* n°36.

structuration du texte est le plus souvent appréhendée au fil de la lecture par des indications verbales et non verbales (Elalouf, 2016) : les enseignants suggèrent des retours à la ligne, entourent des connecteurs qui semblent poser problème (« Et » et « Mais » en tête de phrase notamment), réunissent sous l'étiquette *concordance des temps* les changements dans l'énonciation.

Les éléments ci-dessus conduisent à un deuxième constat qui concerne les stratégies de lecture. Les enseignants « sont massivement plus sensibles aux problèmes locaux qu'aux problèmes globaux » (Halté, 1989 : 20), privilégiant des remarques ayant trait à la phrase plutôt qu'au texte dans son ensemble. Cette tendance peut s'expliquer par la prédominance de la grammaire de phrase dans l'enseignement, rendant les enseignants moins sensibles à l'organisation textuelle. Elalouf (2016), reprenant le protocole de Halté (1989, cf. plus loin), indique de surcroît que la stratégie de lecture majoritaire chez les enseignants est une lecture qui va du local vers le global puisque l'enseignant annote le texte au fur et à mesure qu'il le découvre. Or, la cohérence textuelle s'apprécie à la fois sur les niveaux local (inter et transphrastique) et global (texte).

Le troisième constat porte sur la nature des annotations : la plupart d'entre elles sont verdictives ou injonctives, sans véritable explication, comme le fait remarquer Halté (1989 : 19) :

> « Quatre-vingt pour cent des annotations constatent des manquements de diverses natures à des règles qui ne sont pas explicitées ou, sans même énoncer de constats, imposent une retouche. « Mal dit », « incorrect » sont typiquement verdictifs : ils portent jugement mais ne stipulent pas les motifs ».

Charolles (1978) ajoute que le vocabulaire employé pour dénoncer un manque de cohérence textuelle ne présente souvent aucun caractère technique et traduit plutôt une impression globale de lecture ou une difficulté d'appréhension générale : « incompréhensible », « ne veut rien dire ». En outre, les enseignants rédigent généralement leurs commentaires sous la forme négative et s'expriment plutôt par des abréviations, expressions ou signes divers (notamment « ? ») plutôt que par des phrases complètes, limitant ainsi les pistes d'amélioration et de correction proposées aux élèves.

Outre la mise en évidence des différents types d'annotations des enseignants lors de l'évaluation d'un texte d'élève, cette étude révèle le paradoxe existant dans le mécanisme de l'évaluation d'un texte d'élève. Il y a en effet une contradiction entre l'objectif déclaré de l'enseignant - à savoir aider l'élève à construire des compétences, à améliorer son texte - et les

pratiques mises en œuvre pour y parvenir - des annotations dont la nature ne favorise pas l'objectif à atteindre. Comme l'explique Halté (1989 : 21) :

> « Les enseignants qui ont participé à ce travail […] ne se reconnaissent pas du tout dans le descriptif. L'image qu'il donne les dérange, les heurte même. Ils ne « croient » pas faire ce qu'ils font. Plus : ils ne *veulent* pas le faire ».

En 2014, Elalouf (2016) propose la même expérience que celle d'Halté reposant sur le même texte d'élève, à des enseignants en formation initiale et continue. Elle a donc soumis à nouveau la « copie de Cécile » à des enseignants et analysé leurs annotations. Les constats sont les mêmes trente ans après (Elalouf, 2016 : 9) :

> « Nous avons voulu interroger […] un geste de métier qui résiste aux changements de programmes comme aux travaux en didactique de l'écriture. Parmi les configurations didactiques qui ont sédimenté l'enseignement du français, il frappe par sa résistance, sa capacité à resurgir à tout moment sous forme de modèle didactique en acte [.] »

L'évaluation de textes d'élèves par les enseignants reste donc prioritairement normative, et ce malgré toutes les pistes proposées par la recherche. Cependant, s'il est relativement facile de définir une norme d'un point de vue syntaxique ou grammatical, ce n'est plus le cas lorsqu'il s'agit de définir une norme du point de vue de la cohérence textuelle : « Les phénomènes de cohérence tiennent à l'interprétation et relèvent des domaines pragmatique et sémantique, sans qu'il soit possible de les réduire à des règles d'encodage strictes et univoques » Masseron (2011 : 135). Les problèmes dans l'évaluation soulevés ci-dessus sont caractéristiques d'un « système traditionnel qui, dans les pratiques concernant l'écrit, occulte les problèmes spécifiques de la production de textes au profit de l'acquisition de contenus de savoirs essentiellement morphosyntaxiques ». (Mas, 1989 : 9). Pour faire évoluer ce système, le groupe EVA[39] a construit une grille de critères d'évaluation : *le classement des lieux d'intervention didactique* (CLID) (Mas & *al.*, 1991). Ce document, présenté sous la forme d'un tableau à double entrée reproduit sous la figure 2, permet à l'enseignant de classer les erreurs des élèves et de dresser un portrait des réussites et des échecs en production d'écrit, afin de l'aider à évaluer un texte de manière précise et de ce fait, à transmettre aux élèves des pistes d'amélioration.

[39] groupe Pratiques d'évaluation des écrits

UNITE / NIVEAU	TEXTUEL	RELATIONS ENTRE PHRASES	PHRASTIQUE
PRAGMATIQUE (utilisation des signes en situation)	Prise en compte du destinataire Définition de l'enjeu Type d'écrit	Progression thématique Maîtrise des organisateurs	Organisation thème/rhème Maîtrise des déictiques Maîtrise des déterminants
SEMANTIQUE (relations entre les signes et les « objets » auxquels ils renvoient)	Type de texte Sélection et organisation des informations Cohérence d'ensemble du vocabulaire	Maîtrise des substituts nominaux Respect de la cohérence des informations Maîtrise des connecteurs	Maîtrise du vocabulaire Acceptabilité du sens
MORPHO-SYNTAXIQUE (relations des signes entre eux)	Maîtrise de la structure du texte ou de la séquence Maîtrise des valeurs des temps verbaux	Cohérence dans l'emploi des temps et des modes verbaux Maîtrise des pronoms de reprise	Conjugaison Syntaxe de la phrase Orthographe
ASPECT MATERIEL	Choix du support, de la typographie… Organisation de l'espace page Éventuellement, schémas, illustrations…	Découpage en paragraphes Maîtrise de la ponctuation (délimitation des unités du discours)	Maîtrise de la ponctuation interne à la phrase Utilisation des majuscules Lisibilité graphique

Figure 2 : Grille d'évaluation des écrits CLID, tirée de MAS, M. (1989)

La grille d'évaluation ainsi construite privilégie la grammaire textuelle. Le tableau à double entrée met en rapport l'entrée horizontale des « unités » de découpage d'un texte (le texte entier, les relations entre phrases et la phrase) et l'entrée verticale des « niveaux » de l'analyse (pragmatique, sémantique, morpho-syntaxique, ainsi que l'aspect matériel). Comme l'explique Masseron (2011), le cadre ainsi formé permet de saturer douze cases correspondant à douze questions (ou regroupement de questions) d'évaluation, qui spécifient des critères d'analyse. Le découpage en unités (texte, relations entre phrases et phrase) invite à prendre en

compte à la fois la cohérence (niveau global) et la cohésion (niveau local) du discours. En effet, la cohérence textuelle peut s'appréhender à travers les éléments de la colonne *Textuel*. Par exemple, prise en compte de la situation (qui parle ? à qui ? dans quel but ?) au niveau pragmatique, sélection et organisation des informations au niveau sémantique, concordance des temps au niveau morpho-syntaxique. Les critères de cohésion textuelle se retrouvent quant à eux plutôt dans la colonne *Relations entre phrase* : par exemple, la progression thématique et le statut informationnel au niveau pragmatique, la maîtrise des connecteurs ou celle des substituts nominaux au niveau sémantique, les anaphores au niveau morpho-syntaxique avec la maîtrise des pronoms de reprise. Lorsqu'on « traduit » les critères des colonnes en questions (Masseron, 2011 : 142), on remarque que tous les indicateurs de la colonne *Relations entre phrases* font appel à un jugement de cohérence :

- La cohérence sémantique est-elle assurée (absence de contradiction d'une phrase à l'autre, substituts nominaux appropriés, explicites…) ?
- L'articulation entre phrases ou propositions est-elle marquée efficacement (choix des connecteurs) ?
- La cohérence syntaxique est-elle assurée (utilisation des articles définis, des pronoms de reprise…) ?
- La cohérence temporelle est-elle assurée ?
- La concordance des temps et des modes est-elle respectée ?

La grille d'évaluation des écrits CLID est un des moyens d'évaluer la cohérence et la cohésion textuelles dans les productions écrites des élèves. Une autre proposition de grille d'évaluation fait appel aux 4 méta-règles de la cohérence textuelle définies par Charolles (1978, cf. chapitre 1, section 1.1.). Pour rappel, il s'agit des méta-règles de répétition, progression, non contradiction et relation. Carbonneau et Préfontaine (2005) s'en sont inspirées pour construire une grille d'évaluation contenant des indications de marques adéquates ou inadéquates de cohérence (cf. figure 2 ci-dessous). La grille d'évaluation proposée par les chercheuses québécoises comprend quatre grandes parties, chacune faisant référence à une méta-règle de la cohérence. Chaque grande partie comprend une ou plusieurs composantes (reprise nominale/reprise pronominale pour la partie concernant la première méta-règle par exemple) et chacune de ces composantes comprend plusieurs critères d'évaluation de la cohérence.

GRILLE D'EVALUATION AVEC INDICATION DE MARQUES INADEQUATES ET ADEQUATES DE COHERENCE						
MARQUES INADEQUATES			CRITERES			MARQUES ADEQUATES
META-REGLE DE REPETITION/CONTINUITE						
			Reprise nominale			
			répétition			
			utilisation d'un synonyme			
			utilisation d'un générique			
			utilisation d'une périphrase			
			utilisation d'un déterminant			
			Reprise pronominale			
			référent absent/lien avec le référent			
			référent ambigu			
			genre et nombre			
META-REGLE DE PROGRESSION/ORGANISATION						
			Progression de l'information			
			redondance			
			ellipse/trop d'implicite			
			rythme trop lent			
			Organisation de l'information			
			organisateur textuel			
			découpage en paragraphes			
			regroupement des idées			
META-REGLE DE NON-CONTRADICTION						
			énonciation			
			changement de narrateur			
			paroles rapportées			
			Temps			
			temps du récit			
			temps des verbes			
			Sens			
			sens des mots, des expressions, etc.			

META-REGLE DE RELATION						
			lien entre énoncé et contexte			

Figure 3 : Grille d'évaluation avec indication de marques inadéquates et adéquates de cohérence, tirée de Carbonneau & Préfontaine (2005)

Comme le font remarquer les auteures, cette grille a le mérite de comptabiliser à la fois les marques qui nuisent à la cohérence mais aussi celles qui y contribuent. C'est pour cette raison que les critères d'analyse se trouvent au centre de la grille, laissant la partie de gauche pour inscrire les éléments inadéquats du texte d'élève, c'est-à-dire les éléments qui font obstacle à la cohérence et la partie de droite pour les éléments adéquats, c'est-à-dire ceux qui y contribuent.

Pour utiliser cette grille, les chercheuses proposent de procéder par étapes. Dans un premier temps, les éléments qui nuisent à la cohérence textuelle et ceux qui y concourent sont identifiés dans le texte de l'élève, numérotées, et le numéro est inscrit directement sur la copie de l'élève, comme ci-après.

EXTRAIT D'UN TEXTE D'ELEVE AVEC L'ANNOTATION DES ECARTS A LA COHERENCE

1 2

Un moment donné, Lucien. un jeune homme, revenait de chez sa grand-mère [en voiture].

3 4 5

Chaque fois qu'il venait de chez elle, il prenait toujours le même chemin. Mais à ce moment-là,

6 7 8 9 10

Lucien sortit la tête et regarda une voiture qui n'arrêter pas de le suivre. Lucien panique.

Figure 4 : extrait d'un texte d'élève avec l'annotation des écarts à la cohérence, tiré de Carbonneau et Préfontaine (2005)

Les numéros sont relevés et inscrits dans la grille, dans la case correspondante. En d'autres termes, les marques de cohérence jugées adéquates sont inscrites dans le tableau, dans la colonne à droite des critères d'analyse, par le biais du numéro qui leur a été attribué. Il en va de même avec les marques de cohérence jugées inadéquates, inscrites par leur numéro dans la colonne de gauche, comme dans la figure 3 ci-après.

GRILLE D'EVALUATION AVEC INDICATION DE MARQUES INADEQUATES ET ADEQUATES DE COHERENCE						
MARQUES INADEQUATES			CRITERES	MARQUES ADEQUATES		
META-REGLE DE REPETITION/CONTINUITE						
			Reprise nominale	6	9	
			répétition			
			utilisation d'un synonyme			
			utilisation d'un générique			
			utilisation d'une périphrase			
			utilisation d'un déterminant			
			Reprise pronominale			
			référent absent/lien avec le référent	4		
			référent ambigu			
			genre et nombre			
META-REGLE DE PROGRESSION/ORGANISATION						
			Progression de l'information			
			redondance			
	7	2	ellipse/trop d'implicite			
			rythme trop lent			
			Organisation de l'information			
		5	organisateur textuel	1	3	
			découpage en paragraphes			
			regroupement des idées			

META-REGLE DE NON-CONTRADICTION						
			énonciation			
			changement de narrateur			
			paroles rapportées			
			Temps			
		10	temps du récit			
		8	temps des verbes			
			Sens			
			sens des mots, des expressions, etc.			
META-REGLE DE RELATION						
			lien entre énoncé et contexte			

Note : les chiffres inscrits sur la grille d'évaluation renvoient à ceux inscrits dans l'extrait de texte d'élève (figure 4)

Figure 5 : Grille d'évaluation avec indication de marques inadéquates et adéquates de cohérence, tirée de Carbonneau et Préfontaine (2005)

Par l'utilisation de cette grille, l'élève peut voir d'un seul coup d'œil non seulement les marques de cohérence inadéquates mais également celles qui sont appropriées. C'est un instrument qui permet d'évaluer de manière relativement objective et précise les marques de la cohérence des textes d'élèves. Dans un souci d'efficacité, il semble indispensable toutefois que : « […] cette grille soit conforme aux contenus enseignés, afin que les élèves soient évalués en fonction de leurs connaissances réelles… et qu'ils puissent comprendre le contenu de la grille d'évaluation utilisée par les enseignants. » (*Ibid.* : 81).

Les différents critères d'analyse de la cohérence textuelle élaborés par la recherche font apparaître le recours au jugement de cohérence des enseignants sur les productions de leurs élèves. Or, à l'instar de la cohérence textuelle elle-même, le jugement de cohérence reste difficile à définir.

4.2.2. Les enseignants et le jugement de cohérence

Pour tenter d'expliquer les éléments dont relèvent le jugement de cohérence des enseignants, Rondelli (2008) est partie du constat qu'il existe des degrés de formalisation variables à la base de leurs définitions de la cohérence. Ces définitions relèvent soit des définitions théoriques de la recherche en linguistique, soit des conceptions plus ou moins explicites des enseignants, définitions « non théorisées ou incomplètement théorisées » (Rondelli, 2010a : 55). La chercheuse soulève notamment deux questions pour explorer les représentations de la notion de cohérence textuelle des enseignants.

La première interrogation porte sur les critères sur lesquels les professeurs des écoles s'appuient pour juger la cohérence d'un texte d'élève. Pour répondre à cette question, Rondelli (2008 ; 2010a) a comparé les définitions savantes et les représentations « proto-théoriques » des enseignants. Dans ces travaux, elle propose à des élèves de cycle 3 d'écrire la suite d'un texte narratif (325 textes). Ces textes sont ensuite évalués par des enseignants, qui ont pour seule consigne « La suite écrite par l'élève est-elle réussie ou non ? Vous pouvez annoter le texte. ». Les annotations et les commentaires sont étudiés. Il ressort de ces travaux plusieurs facteurs qui influencent le jugement de cohérence des enseignants. En premier lieu, les enseignants jugent la cohérence d'un texte d'abord en fonction des indicateurs linguistiques de l'emploi des temps et de référenciation. L'emploi des temps est jugé réussi si l'articulation entre passé simple et imparfait est homogène, « ce qui provoque une mise en relief clairement identifiable » (Rondelli, 2010a : 65). Les anaphores qui participent à la cohérence du texte, selon les enseignants, sont celles où l'accessibilité référentielle est claire, c'est-à-dire qu'il n'y a pas d'ambiguïté référentielle. À titre d'exemple, Rondelli (2010a) cite l'extrait de texte d'élève suivant, qui présente un cas d'ambiguïté bien connu (cf. chapitre 3, section 3.2.) :

(118) Jarod prit un bâton et tapait sur le requin. Mais celui-ci cassa le bâton.

Dans cet extrait, plusieurs enseignants ont sanctionné l'emploi du pronom *celui-ci* dans la deuxième phrase, arguant que le référent du pronom n'est pas clair : est-ce Jarod ou le requin ? En effet, l'ambiguïté tient au fait que les deux référents potentiels sont deux noms au masculin singulier, il n'y a donc pas de moyens grammaticaux de choisir la bonne source. En outre, le verbe *casser* peut être employé aussi bien pour un référent humain comme *Jarod* que non humain comme *le requin*.

Un autre facteur entrant en jeu dans le jugement de cohérence des enseignants est celui de la cohérence sémantique. Comme l'explique Rondelli (2010a : 70) : « le lecteur construit une signification dans une dynamique fondée sur une sélection thématique qu'il réalise à partir des scénarios intertextuels qu'il connait ». Les mots et leurs entours choisis par les élèves ont un rôle de marqueur de topic et ouvrent un ou des monde(s). Pour qu'un texte soit jugé cohérent, il faut que le(s) monde(s) construit(s) par les mots clés saillants soi(en)t en adéquation avec le monde du texte-amorce et le monde réel.

La deuxième interrogation que relève Rondelli porte sur les accords ou les écarts entre les critères mobilisés par les enseignants et les descriptions linguistiques. D'après les travaux de la chercheuse (Rondelli, 2010a ; 2010b), il semblerait que les enseignants ont une vision plutôt restrictive des liages sémantiques (cohésion verbale) et référentiels (anaphores). En effet, en ce qui concerne la cohésion verbale, les enseignants ont tendance à demander à l'élève une reprise systématique des temps du texte amorce, car toute alternance temporelle est interprétée comme une rupture. En ce sens, ils rejoignent la définition de Combettes (1995) pour qui le critère d'homogénéité énonciative est déterminant pour juger de la bonne cohésion (verbale) d'un texte. En ce qui concerne les anaphores, la majorité des enseignants détectent les ambiguïtés référentielles et sanctionnent les marques de la référence. Pour eux, la cohésion nominale est un élément important car elle établit un lien étroit entre un référent et sa reprise. Pour autant, il semblerait que les enseignants tiennent compte uniquement du cotexte immédiat et de gauche pour résoudre la référence, faisant ainsi preuve d'une conception « antécédentiste » (Apothéloz, 1995, cité par Rondelli, 2010a : 74 ; cf. chapitre 2 de notre recherche). Cette conception met ainsi de côté la question de l'environnement discursif dans l'interprétation des anaphores. En effet, la recherche du référent est, pour la majorité d'entre eux, strictement locale et grammaticale, s'appuyant sur des critères de proximité textuelle et de continuité linguistique au niveau du genre et du nombre.

Un autre point sur lequel les définitions des enseignants et celles de la linguistique sont divergentes concerne les *effets de monde*. En effet, chaque discours ouvre un univers (ou monde) fictionnel ou non, dans lequel prennent place les faits relatés[40] (Charolles, 1997). Dans les travaux de Rondelli par exemple, le texte amorce que les élèves doivent compléter ouvre un monde spécifique dans lequel prennent place un certain décor et des personnages, en l'occurrence un radeau sur la mer et des nageurs. Les enseignants sont partagés quant à

[40] Pour plus de détail sur cette question, nous renvoyons à toute la littérature concernant les cadres/univers de discours et les questions d'ancrage discursif.

l'évaluation de cet aspect de la cohérence. Pour certains, le(s) monde(s) construit(s) doi(ven)t rendre cohérente l'histoire en vue de sa résolution et de ce fait être en adéquation avec le monde ouvert par le texte amorce. En d'autres termes, ces enseignants sanctionneraient une suite de texte dans laquelle l'élève aurait ouvert un univers comprenant la lune et des astronautes par exemple, n'ayant *a priori* aucun rapport avec le radeau en mer et les nageurs du texte amorce. Le critère de l'intelligibilité est ainsi mis en avant, privilégiant une certaine homogénéité thématique et un respect du vraisemblable qui « s'appui[e] sur le savoir humain acquis ou considéré comme « vrai » dans le cadre de leur [les enseignants] propre encyclopédie » (Rondelli, 2010a : 75). Pour d'autres enseignants en revanche, la réception esthétique est aussi importante : le texte de l'élève peut s'affranchir du genre prédéfini par le texte amorce et interagir avec d'autres genres.

À travers ces travaux, il apparaît clairement que les enseignants s'appuient sur des critères linguistiques pour juger de la cohérence d'un texte. Cependant, ces critères sont incomplets au vu des définitions savantes de la cohérence, et parfois même, montrent des variations inter-individuelles en fonction des évaluateurs.

Toutefois, on ne peut reprocher aux enseignants, néophytes en linguistique, d'être incomplets là où les spécialistes eux-mêmes ne s'accordent pas sur les définitions de cohérence et de cohésion. En effet, nous avons montré, dans le premier chapitre de cette thèse, que ces notions restent complexes à définir. Rondelli (2008 : 583) conclut ainsi que :

> « le jugement de cohérence, loin de se concentrer sur la liste des éléments morphosyntaxiques déclinés dans les programmes, s'appuie sur des aspects sémantiques et des procédures psychologiques que les enseignants mettent en œuvre de façon implicite ».

Il est donc difficile de mettre en œuvre des critères précis et relativement objectifs pour juger de la cohérence d'un texte.

SYNTHESE des notions centrales du chapitre 4

À travers ce chapitre, nous avons abordé la référence sous un angle didactique. En nous interrogeant sur l'enseignement des marques de cohérence textuelle, nous avons dressé un bref historique de l'apparition des notions de cohérence et cohésion textuelles dans les

programmes d'enseignement. Les notions de cohérence et cohésion textuelles, ainsi que les marques anaphoriques, sont apparues pour la première fois dans les préconisations de 1992. Même si depuis cette date, la définition de ces notions dans les programmes a connu quelques fluctuations en terme de rigueur scientifique, elles semblent s'inscrire durablement dans les savoirs à enseigner à l'école.

Cette vue d'ensemble a permis de montrer que l'enseignement de la grammaire à l'école adopte majoritairement une approche phrastique, s'attachant à décrire les unités de la phrase mais ne tenant pas compte de la globalité de l'énoncé (Combettes & Charolles, 1999 ; Combettes, 2016). Nous avons pu également constater comment les apports des théories linguistiques et psycholinguistiques de l'analyse du discours (Benveniste, 1974 ; Halliday & Hasan, 1976 ; Van Dijk & Kintsch, 1983), donnant lieu à une approche textuelle de la grammaire scolaire, ont été diffusés dans l'enseignement et ont pu apparaître dans les programmes.

Malgré le fait que la cohérence et la cohésion textuelles soient des notions au programme depuis plusieurs années, nous avons remarqué que l'enseignement des marqueurs référentiels est souvent source de difficultés (Reichler-Béguelin, 1988b). D'abord pour les élèves qui ne sont pas suffisamment instruits sur ce point, mais aussi pour les enseignants qui sont confrontés très souvent à un manque d'outils didactiques et pédagogiques sur la question (Bergeron & Blain, 2003). Nous avons donc exposé d'abord les objectifs principaux à atteindre dans l'apprentissage de la référence : dépassement du traitement local de la référence au profit d'un traitement global sur l'ensemble du texte, gestion des remaniements syntaxiques (transformations, condensation, déplacement), prise en compte de la diversité des substituts possibles et de leur alternance tout au long du texte (Garcia-Debanc, 1993 ; Reichler-Beguelin, 1993). Ensuite nous avons pu proposer quelques situations, tirées des recherches linguistiques et didactiques, afin d'aider les enseignants dans cette tâche (Tauveron, 1987; Charolles, 1988c ; Masseron, 1989 ; Djebbour & Lartigue, 1994 ; Schnedecker, 1995).

Tout savoir enseigné devant aussi faire l'objet d'une évaluation, nous avons interrogé la place de la cohésion textuelle dans l'évaluation. Le jugement de cohérence que portent les enseignants sur les textes d'élèves dépend de savoirs linguistiques acquis mais incomplets au vu des définitions savantes, et souvent rapporté au respect de la norme grammaticale (Halté, 1989 ; Masseron, 2011 ; Elalouf, 2016). En outre, des connaissances sémantiques et

psychologiques implicites entrent en jeu et peuvent varier d'un individu à l'autre (Rondelli 2008, 2010a, 2010b). Le jugement de cohérence a donc une part de subjectivité, ce qui conforte l'idée exprimée dans le premier chapitre de nos travaux que la cohérence est une notion bien difficile à décrire et manipuler. Pour cette raison, nous avons proposé des critères d'évaluation de la cohérence textuelle recensés dans différents travaux (Carbonneau & Préfontaine, 2005 ; Masseron, 2011).

Ce chapitre clôt ainsi la première partie de notre étude. La cohérence textuelle dans les productions écrites des élèves étant au cœur de notre recherche, nous appuyons notre analyse sur un corpus de textes d'élèves. Dans ce contexte, il nous paraît nécessaire de dresser un rapide état de la question des corpus de textes scolaires. C'est ce que nous nous proposons de faire dans le premier chapitre de la deuxième partie de la thèse (chapitre 5).

PARTIE II :
Mise en œuvre de l'étude et méthodologie

Dans les chapitres précédents (1 à 4), nous avons présenté le cadre théorique de notre étude et montré les enjeux à la fois linguistiques, psycholinguistiques et didactiques de notre recherche. Nous consacrons la deuxième partie de notre thèse à la présentation de la mise en œuvre de notre étude et des aspects méthodologiques qui ont présidé à la constitution et à l'analyse de notre corpus, à travers trois chapitres.

Le premier (chapitre 5) présente un état des lieux non exhaustif des méthodologies utilisées pour la constitution de corpus de textes d'élèves. Le deuxième (chapitre 6) expose la nos questions de recherche et indique la méthodologie employée pour le recueil de nos données et la constitution du corpus de l'étude. Le troisième (chapitre 7) détaille les critères d'annotation, c'est-à-dire les indices annotés tirés de nos hypothèses, que nous avons mis en place pour l'analyse quantitative et qualitative de notre corpus d'étude, ainsi que les problèmes d'interprétation rencontrés.

CHAPITRE 5 : De l'analyse de textes d'élèves à la constitution de grands corpus de textes scolaires

Lorsque l'idée de constituer des corpus de textes scolaires a émergé, dans les années 1990, elle était aussi novatrice que celle de constituer des corpus de français parlé quelques décennies plus tôt (voir les travaux précurseurs de Blanche-Benveniste notamment). De même que pendant longtemps, les linguistes ne s'étaient pas intéressés à l'oral, le français écrit des élèves n'éveillait pas (ou peu) l'intérêt des linguistes (Elalouf & Boré, 2007 ; Roubaud, 2017). Comme le souligne Roubaud (2017), les chercheurs qui se préoccupaient de cette question se retrouvaient confrontés au même enjeu qui avait présidé à l'émergence de l'étude du français parlé : donner une légitimité aux données langagières étudiées avec la constitution d'un grand corpus de textes scolaires.

Il nous semble judicieux de préciser un point de terminologie ici. Dans son acception générale, un corpus est un recueil de pièces ou de documents qui concernent une même matière, discipline ou doctrine (Dalbera, 2002). En linguistique, un corpus est « une collection de données langagières qui sont sélectionnées et organisées selon des critères linguistiques explicites pour servir d'échantillon du langage » (Sinclair, 1996, cité par Habert & *al.,* 1997 : 11).

De ce fait, en 5.1., nous allons voir que les premiers corpus de productions écrites d'apprenants en français langue première étaient des échantillons limités de textes d'élèves, constitués pour l'analyse d'un fait de langue précis. Depuis quelques années, un changement d'échelle progressif est constaté : on passe de l'analyse de quelques textes à la constitution de grands corpus de textes scolaires. Ce changement d'échelle a été rendu possible par la volonté de mettre à la disposition de la communauté scientifique des banques de textes servant d'appui à une analyse croisée des phénomènes linguistiques.

Toutefois, en 5.2., nous montrons que quelle que soit la taille des corpus de textes d'apprenants, leurs caractéristiques soulèvent de nombreuses difficultés, notamment d'ordre méthodologique, qui nécessitent souvent de faire des choix.

Enfin, en 5.3., nous nous intéressons plus particulièrement à l'exploitation des corpus, à travers notamment une description succincte de quelques outils d'aide à l'annotation utilisés pour l'exploitation de corpus de textes scolaires. Nous terminons par le constat qu'il n'existe

pas (encore) de corpus de textes scolaires dont l'objectif soit l'étude de la cohésion textuelle, et plus particulièrement des chaînes de référence.

5.1. Les corpus de productions écrites d'apprenants en français langue première : vers une meilleure représentativité

Les écrits d'élèves suscitent de nos jours un intérêt croissant dans de nombreux domaines des sciences du langage (linguistique, psycholinguistique, sociolinguistique) et particulièrement en didactique du français langue première. Comme l'expliquent Doquet & *al.* (2017a : 7) :

> « cet intérêt s'explique à la fois par la singularité de l'objet discursif qu'ils constituent et par la rareté des études empiriques appuyées sur des ensembles de textes de grande envergure ».

En règle générale, les grands corpus visent à devenir des corpus de référence, c'est-à-dire représentatifs de la norme prescrite (Bowker & Pearson, 2002, cités par Lopez, 2010). En d'autres termes, les grands corpus ont pour objectif de constituer un échantillon de données suffisamment grand pour en déduire des règles applicables à toute une population. Toutefois, la représentativité ainsi visée n'est qu'un idéal qu'il est difficile d'atteindre. En effet, les variables à introduire pour y parvenir sont nombreuses et concernent plusieurs niveaux parmi lesquels milieu sociologique, performances scolaires, genres textuels étudiés.

Aussi, en 5.1.1., nous exposons quelques-unes des limites qui expliquent le développement récent de corpus écrits d'apprenants en français langue première. En 5.1.2. nous indiquons comment la prise en compte des enjeux à la fois scientifiques et sociétaux de tels types de corpus a amené plusieurs tentatives de constitution de grands corpus d'écrits scolaires.

5.1.1. Les enjeux de la constitution des corpus d'écrits scolaires

Les corpus étudiés en linguistique de corpus sont nombreux, variés et les grands corpus de référence sont fréquents. Pour l'anglais par exemple, on peut citer le *British*

National Corpus (*BNC*) et ses 100 millions de mots ou le *Corpus of Contemporary American English* (*COCA*) avec ses 520 millions de mots. Pour la langue française, *Frantext* par exemple, a servi de base à l'élaboration du *Trésor de la Langue Française* (*TLF*). Toutefois, les grands corpus réunissant des productions d'apprenants sont plutôt récents (1980/1990) et beaucoup moins nombreux.

Pour l'anglais par exemple, les corpus d'apprenants sont plus nombreux que dans les autres langues. Toutefois, comme le fait remarquer Granger (2009), il s'agit principalement de corpus d'apprenants en langue seconde : les données sont recueillies auprès de locuteurs adultes ou étudiants apprenant l'anglais mais dont ce n'est pas la langue maternelle (par exemple, l'*International Corpus of Learner English (ICLE),* le *Hong Kong University of Science and Technology Learner Corpus*). En anglais langue première, le premier grand corpus d'écrits scolaires, le *Lancaster Corpus of Children's Project Writing*[41], a pris forme entre 1996 et 2000 (Wolfarth & *al.*, 2017). Ce corpus, constitué dans l'objectif d'explorer la diversité des écrits des élèves, rassemble des textes produits par des élèves de 8 à 12 ans dans 37 écoles. En ce qui concerne le français, le nombre de grands corpus est bien moindre et le constat est le même que pour l'anglais : les grands corpus de textes scolaires sont mieux représentés en Français Langue Etrangère qu'en français langue première (Granger 2009 ; Bilger & Cappeau, 2013 ; Wolfarth & *al.*, 2017).

La rareté des grands corpus d'écrits scolaires en français langue première tient au fait que : « la tradition des travaux en didactique du français langue première sur l'analyse des textes scolaires […] s'attache généralement à une analyse fine d'un échantillon très limité de textes (Fabre-Cols, 2000 ; Cappeau & Roubaud, 2005 ; Boré, 2011 ; Masseron, 2005) » comme l'indiquent Garcia-Debanc & *al.* (2017 : 158). En effet, plusieurs autres chercheurs constatent également que les corpus conséquents sont rares dans les travaux sur l'écriture scolaire (David, 2000a ; Elalouf, 2004 ; Elalouf & Boré, 2007 ; Roubaud, 2017).

Plusieurs facteurs peuvent expliquer cette rareté. L'accès aux classes est rendu plus ou moins difficile en fonction des politiques des établissements. D'une part, il nécessite des autorisations administratives pour qui n'est pas en relation directe avec le milieu scolaire. D'autre part, l'enseignant peut redouter d'offrir sa classe et ses méthodes d'enseignement au regard du chercheur, cet autre au fond de la classe dont le regard d'expert peut le mettre mal à

[41] http://www.lancaster.ac.uk/fass/projects/lever/index.htm.

l'aise. Des contraintes purement pratiques entrent également en ligne de compte. En effet, le recueil de données linguistiques s'avère complexe. La collecte de données dans les classes est également coûteuse en temps : le recueil exige disponibilité et précautions déontologiques (Elalouf, 2011).

La rareté des études d'écrits scolaires tient aussi au fait que leurs spécificités en font des textes difficiles à analyser (cf. section 4.2.) : l'écart à la norme notamment les rend « réfractaires à la numérisation et à l'étiquetage » (Elalouf & Boré, 2007 : 53). Pourtant, des corpus importants ont pu être publiés, notamment dans de nombreuses thèses en sciences du langage relevant de la didactique du français langue première. Toutefois, l'analyse des textes qui les composent reste partielle et leur accès difficile (Doquet, David & Fleury, 2017a) car « l'interprétation des textes dépend étroitement des moyens dont dispose le chercheur pour accéder à leur intelligibilité » (Elalouf, 2011 : 59), et leur présentation doit rendre compte à la fois d'évolutions individuelles (échelle de l'élève) et de dynamiques collectives (échelle de la classe).

La rareté des grands corpus d'écrits scolaires dans les travaux en didactique est d'autant plus paradoxale que c'est grâce à l'analyse de tels types de corpus qu'ont été réalisées de grandes avancées en didactique de l'écriture, comme le montrent Elalouf & Boré (2007), qu'il s'agisse de l'étude des brouillons d'élèves (Fabre-Cols, 1990, 2002), de l'analyse de l'évolution des textes scolaires selon les dispositifs didactiques (Garcia-Debanc, 1990), de la relation entre modalités d'écriture et choix des supports (Plane, 1995 ; 2000), de l'étude des procédés d'écriture (Bonnet, 1994 ; Bonnet & *al.*, 1998) ou de l'inscription du sujet scripteur dans ses textes (Chabanne & Bucheton, 2002) .

Il nous faut préciser ici un point d'analyse. En effet, nous devons relever qu'à notre connaissance, la littérature scientifique ne fait nulle part état d'une définition nette de ce qu'est un grand corpus de textes scolaires. Par conséquent, nous pouvons nous interroger : à partir de combien de textes d'élèves peut-on considérer un corpus comme significatif ? La significativité d'un corpus de textes scolaires doit-elle d'ailleurs s'appréhender en nombre de textes ou en nombre de mots ? Cette dernière question est d'autant plus complexe lorsqu'on connaît toutes les difficultés méthodologiques que soulèvent les écrits de ce type. Sur les plans de l'orthographe et de la segmentation, les textes d'élèves sont peu ou non normés dans la mesure où ils émanent d'apprentis scripteur. Si l'on décide que la significativité d'un grand corpus de textes scolaires s'appréhende par le nombre de mots, à la manière des corpus de

référence, l'aspect non normé de la segmentation peut donc devenir un obstacle au comptage en nombre de mots.

De surcroît, si en linguistique de corpus, on parle de grand corpus ou de corpus de référence lorsqu'il s'agit de corpus contenant plusieurs centaines de millions de mots, il est impossible (ou à tout le moins extrêmement complexe) d'en rassembler autant lorsqu'il s'agit d'écrits d'élèves. Doquet & *al.* (2017b : 136), avec les 2300 textes (textes et avant-textes) du « grand » corpus ECRISCOL n'atteignent que 35 000 mots environ. Cependant, le critère de la taille n'est peut-être pas le plus important, comme le souligne Ragan (2001, cité par Granger, 2009 : 262) :

> « the size of the sample is less important than the preparation and tailoring of the language product and its subsequent corpus application to draw attention to an individual or group profile of learner language use." [42].

L'examen des travaux ayant conduit à des avancées en matière de didactique de l'écriture met en évidence une autre limite que l'on trouve aux corpus d'écrits d'apprenants en français langue première. En effet, s'il s'agit des quelques rares études s'appuyant sur de « grands » corpus d'écrits scolaires, ces travaux analysent un seul fait de langue à la fois, réduisant ainsi le champ de recherche, comme l'expliquent Gunnarsson-Largy & Auriac-Slusarczyk (2013 : 8) :

> « Le chercheur suivant une orientation scientifique oriente ainsi le regard de l'enseignant ou de l'étudiant sur l'aspect scientifique à partir duquel il renouvelle les connaissances scientifiques ».

La plupart du temps, le corpus constitué a pour objectif de rendre compte du mieux possible du fait de langue qui va être étudié, comme par exemple la ponctuation, l'orthographe, certaines relations grammaticales. Pour cette raison, sont étudiés uniquement les textes répondant aux critères d'analyse définis au préalable. De cette manière, beaucoup d'informations jugées non pertinentes par rapport aux critères choisis sont laissées de côté : par exemple un premier jet, une réécriture ou un texte jugé trop raturé pour être utile à autre chose qu'une étude sur les ratures. Or, le croisement de différents faits de langue peut mettre en lumière des éléments nouveaux. De la même manière, le croisement de différents phénomènes (phonologie, syntaxe, sémantique, etc., mais aussi linguistique, psycholinguistique, didactique, etc.) peut être très productif (Auriac-Slusarczyk & *al.*, 2008).

[42] « la taille de l'échantillon est moins importante que la préparation et la confection du matériau langagier et de ses applications en corpus pour attirer l'attention sur le profil d'un individu ou d'un groupe d'apprenants. » (notre traduction)

Plus particulièrement, dans le cadre de l'enseignement/apprentissage de l'écriture à l'école, un corpus de textes scolaires permet de mener à la fois une analyse linguistique des productions recueillies et une analyse didactique des dispositifs conçus pour l'enseignement et l'apprentissage de l'écriture, depuis la graphie et la segmentation des énoncés jusqu'à la production de textes dans leurs différentes étapes (Doquet, David & Fleury, 2017a : 32). Cette double démarche est d'ailleurs préconisée par Masseron (2011 : 129) :

> « il apparaît légitime d'attendre qu'une analyse linguistique, caractérisée par une méthode et des objets de recherche identifiés, « interagisse avec » une démarche en didactique […] avec l'objectif commun d'interroger des performances écrites ».

En effet, comme l'indique l'auteur (*ibid.*), la confrontation d'une analyse linguistique et d'une démarche didactique peut construire ou renouveler des usages langagiers des élèves, que ce soit à un moment donné de l'apprentissage ou bien de manière longitudinale (Masseron, 2011). D'autre part, les écrits scolaires soulevant des problèmes linguistiques gagneraient à être confrontés à une analyse qui prend comme point de vue à la fois la linguistique et d'autres domaines d'étude.

La complémentarité des approches donnerait également aux grands corpus d'écrits scolaires un objectif de formation. En effet, l'appui sur des textes d'élèves pour enseigner des notions, telles que la grammaire par exemple (cf. chapitre 3, section 3.2.), reste rare « si l'on excepte l'utilisation traditionnelle de ces textes comme réservoirs à erreurs » (Doquet, 2013 : 120). Or, comme l'indiquent Boré & Elalouf (2017 : 32), en devenant une banque de données composée de textes dans leurs différents états, représentatifs du développement des compétences sur plusieurs cycles, les grands corpus scolaires permettent :

> « […] au formateur – et à travers lui à l'enseignant – de se constituer une culture de ces textes qui ne prennent souvent du relief que par contraste avec d'autres, ce qui oblige à s'interroger sur les effets d'un dispositif d'enseignement-apprentissage sur chaque texte dans sa singularité et sur la classe dans son ensemble. »

Aussi, la constitution de grands corpus d'écrits scolaires correspond à deux enjeux majeurs (Elalouf, 2011 ; Garcia-Debanc & *al.*, 2017 ; Jacques & Rinck, 2017). Le premier est un enjeu scientifique : la progression des connaissances sur la mise en place de compétences rédactionnelles correspondant à différents niveaux d'analyse et à différentes phases de l'apprentissage. En effet, la notion de progression, « si présente dans les pratiques d'enseignement » est « relativement mise en veilleuse sur le plan théorique » et pourtant elle

est « cruciale » (Nonnon, 2010). Brissaud & Grossmann (2009) font d'ailleurs remarquer qu'il n'existe pas encore « d'image panoramique de ce que serait l'évolution de la compétence grammaticale de 6 à 16 ans » et le constat est le même en ce qui concerne les compétences rédactionnelles et lexicales.

D'autre part, le deuxième enjeu majeur de la constitution de grands corpus d'écrits scolaires est un enjeu sociétal : la constitution de ressources pour les chercheurs, les formateurs d'enseignants, les enseignants et les étudiants futurs enseignants.

5.1.2. Les quelques tentatives de constitution de grands corpus d'écrits scolaires en français

Plusieurs tentatives de constitution de grands corpus d'écrits scolaires ont vu le jour depuis les années 2000, le nombre de textes les composant augmentant significativement avec le temps. Nous nous proposons ici d'en faire une rapide revue qui ne prétend pas à l'exhaustivité.

Dans le domaine de la constitution de grands corpus, les précurseurs Elalouf & Boré (Elalouf & *al*., 2005 ; Elalouf & Boré, 2007), qui présentent pour la première fois une tentative de diffusion d'un corpus d'écrits d'élèves permettant d'observer un certain nombre de faits de langue et de discours de l'écrit entre 10 et 14 ans, *via* un CD-ROM. Le corpus publié présente l'ensemble des textes produits dans leurs différentes versions, pour chaque activité d'écriture (soit au total 728 textes). Dans ce projet, les chercheurs se sont appuyés sur des situations écologiques en observant une séquence d'apprentissage constituée de plusieurs séances reliées par un objectif d'apprentissage commun. Ainsi, les textes produits dans la séquence sont mis en relation avec les textes lus en classe, mais aussi avec les référents culturels explicitement convoqués ou simplement évoqués dans la situation d'écriture. Les textes produits sont également mis en relation avec les commentaires et interventions orales ou écrites de l'enseignant et/ou des pairs sur les textes (Boré & Elalouf, 2017). L'objectif premier de ce corpus était d'analyser les écrits en relation avec leur contexte de production. Toutefois, le fait que le corpus constitué soit mis à disposition de la communauté rend possibles des travaux ultérieurs sur des sujets aussi variés qu'une analyse selon différents

angles, par exemple la comparaison d'un sous-corpus, du niveau d'enseignement, du genre ou du dispositif d'écriture.

À la suite de ces travaux précurseurs, un groupement de chercheurs, le GDR sur la Production Verbale Ecrite[43], s'est donné pour objectif un « travail d'étude approfondie d'écrits d'élèves sur la base d'un corpus textuel commun, construit pour permettre, à terme, à des enseignants d'en tirer parti. » (Auriac-Slusarczyk & *al.*, 2008). S'appuyant sur ce corpus, les chercheurs ont étudié le rapport à l'écriture dans le développement des compétences du jeune scripteur en faisant varier notamment deux genres textuels : le récit et le compte-rendu. Ils se sont demandés quelles variations induisent ces deux genres dans la manière dont les élèves vont traiter la langue et s'il se dégage des invariants, susceptibles de mettre en évidence des difficultés types ou récurrentes chez de jeunes élèves (Gunnarsson-Largy, & Auriac-Slusarczyk, 2013). Pour répondre à ces questions, il est proposé à des élèves de 9 à 14 ans une tâche d'écriture suivi d'une réécriture. Pour le récit, les élèves doivent raconter une histoire en s'appuyant sur six images issues d'un album de jeunesse dont une grenouille est l'héroïne. Pour le compte-rendu, les élèves doivent écrire un texte pour présenter le cycle de vie de la grenouille, en s'appuyant sur six images issues d'un manuel de sciences. Ces supports d'écriture ont d'ailleurs participé à la dénomination de ce travail comme *corpus Grenouille*.

L'objectif de ce travail est la constitution d'un corpus de textes d'élèves (plus de 400) dont la mise à disposition permette une meilleure vue d'ensemble des faits de langue étudiés, car, comme le résument les auteurs (Auriac-Slusarczyk & *al.*, 2008) :

> « il semble plus facile de comprendre comment les élèves produisent un texte, lorsqu'est auscultée une tâche identique, plutôt de que disposer d'études scientifiques isolées, qui montrent de manière alors plus parcellaire les mêmes éléments ».

En effet, le corpus Grenouille présente un intérêt méthodologique car il permet : « des observations croisées, des recoupements soit dans une même classe soit entre les classes, de dégager des régularités ou d'envisager des tendances » (Cappeau, 2013 : 42).

Plus récemment, dans le cadre d'une opération de recherche intitulée *Analyse linguistique de l'écriture scolaire, ECRISCOL*[44], le laboratoire Clesthia[45] a commencé en

[43] GDR-Production Verbale Ecrite-CNRS 2657
[44] http///www.univ-paris3.fr/ecriscol

2013 à constituer un grand corpus de textes scolaires, dont la taille et l'étendue en âge visent la représentativité (David & Doquet, 2016 ; Doquet & *al.*, 2017b). En effet, les chercheurs ont recueilli à ce jour 1225 textes et les avant-textes correspondants (notes, brouillons…). Les écrits ont été réalisés à différents âges de la scolarité, correspondants aux trois paliers du socle commun de connaissances, de compétences et de culture, ainsi qu'à la fin de la scolarité secondaire : au CE1 (7-8 ans), au CM2 (10-11 ans), en 3°/début 2nde (14-15 ans), et en terminale/entrée à l'université (17-18 ans). Pour rester au plus près de la réalité des textes produits dans le cadre scolaire, les chercheurs ont fait le choix d'un corpus écologique, c'est-à-dire : « d'aller chercher dans les classes des écrits tels que le quotidien de l'enseignement permet de les faire émerger » (Doquet & *al.*, 2017b : 136). Ils ont donc proposé aux élèves une tâche courante dans la scolarité : écrire la suite d'un récit. Toutefois, les enseignants restent libres de choisir le texte narratif amorce et les conditions didactiques afin que celles-ci puissent constituer des variables d'étude, dans la complémentarité d'une étude à la fois linguistique et didactique caractéristique des grands corpus de textes scolaires, comme nous l'avons vu *supra* (cf. section 4.1.1.). Comme l'expliquent Doquet & *al.*, (2017b : 134), l'objectif de ce corpus est de :

> « de recueillir un grand nombre d'écrits variés pour en analyser les caractéristiques linguistiques et discursives […] afin de tracer une cartographie des compétences selon les différentes variables (socio-culturelles, didactiques, identitaires) prises en compte. »

Encore plus récemment, le projet Scoledit, porté par le LIDILEM[46], vise un autre changement d'échelle puisqu'il a pour objectif de réunir à terme plusieurs milliers de productions d'élèves (Wolfarth & *al.*, 2017). Ce projet s'appuie sur le recueil de données provenant de 131 classes de CP (2507 textes) effectué dans le cadre du projet national IFE *Lire-Ecrire au CP* [47]. Pour le CP, toutes les classes ont un protocole expérimental commun : les élèves doivent raconter l'histoire d'un petit chat en s'appuyant sur quatre images présentant ses aventures. Le projet Scoledit, dans l'objectif d'une étude longitudinale, recueille les textes des mêmes élèves chaque année au mois de juin, à partir du CP (juin 2014) jusqu'à leur arrivée en CM2 (juin 2018). L'objectif de la constitution de ce grand corpus est

---

[45] Clesthia – Langage, Systèmes, Discours. EA 7345. Université Paris 3.

[46] LIDILEM – Laboratoire de Linguistique et Didactique des Langues Etrangères et Maternelles, EA 609. Université de Grenoble Alpes.

[47] *Lire – Ecrire au CP*, projet coordonné par Roland Goigoux, financé par la DGESCO, l'IFé, et le laboratoire de recherche Acté (Clermond-Ferrand).

de « permettre de rendre compte des évolutions des procédés d'écriture à différents moments de la scolarisation de l'école primaire » (Wolfarth & *al.*, 2017 : 187).

Les travaux exposés ici, que ce soit les corpus déjà constitués ou en cours de constitution, ont tous en commun l'objectif d'explorer et décrire les écrits des élèves. En effet, « réunir des corpus représentatifs, comparables, cumulables est un enjeu essentiel pour la production de connaissances en didactique du français. » (Elalouf, 2011 : 57). Cependant, pour réussir à exploiter au mieux ces corpus d'écrits d'apprenants, « l'outillage de l'analyse de corpus, confronté à la diffusion régulière d'écrits non standards, doit pouvoir s'appuyer sur des modules d'analyse adaptés aux écarts à la norme qui spécifient les différents corpus » (Doquet & *al.* 2017b : 152). En effet, les corpus d'écrits scolaires sont caractérisés par leur écart à la norme, rendant leur analyse outillée difficile. Toutefois, nous allons le voir, ce n'est pas la seule difficulté propre à ce type de corpus.

5.2. Quelques difficultés pour la constitution et le traitement de corpus d'écrits scolaires

Travailler *avec* et *sur* des corpus d'écrits scolaires soulève de nombreuses difficultés. En 5.2.1., nous exposons les questions d'ordre épistémologique et méthodologique qu'engendre l'exploitation de ce type de corpus. Pour contourner la difficulté que représente le manque de normalisation des écrits d'élèves, il faut passer par une étape de transcription des textes. En 5.2.2., nous voyons que cette étape est loin d'être simple et constitue encore une difficulté dans l'exploitation des corpus. Une fois toutes les difficultés résolues, il reste encore à annoter les textes afin de les analyser de manière plus fine. En 5.2.3., nous nous interrogeons sur le type d'outil d'aide à l'annotation à utiliser pour cette phase d'exploitation du corpus.

5.2.1. Des difficultés épistémologiques et méthodologiques

La première difficulté dans la constitution d'un corpus d'écrits scolaires est sa classification en genres d'écrits scolaires. En effet, la notion de *corpus* renvoie à un ensemble, à une collection (cf. introduction de ce chapitre). Dans les corpus de textes scolaires, la notion de corpus renvoie à un ensemble de textes présentant une certaine unité de genre (Dalbera, 2002). Boré (2007) montre la complexité des notions convoquées lorsqu'il s'agit d'un corpus de genres d'écrits scolaires. Le terme *scolaire* désigne l'institution de laquelle procèdent ces genres. L'expression « genres scolaires *écrits* » peut désigner à la fois les genres prescrits, lus et produits à l'école. La complexité d'une définition des genres enseignés *à* et *par* l'école est donc grande. Aussi, les textes scolaires – « entendus comme des productions appartenant à des genres écrits prescrits par l'école » (Boré, 2007 : 20) – pourraient être considérés comme un ensemble de textes relativement homogènes. En effet, il semblerait normal d'envisager qu'un conte ou un récit policier produit par des élèves d'une même tranche d'âge présente des ressemblances de structure et de langue. Or, Boré & Elalouf (2017 : 41) rappellent qu'« il n'est pas sûr que l'on puisse comparer des productions fictionnelles appartenant à des sous-genres vraiment distincts (récits policiers, contes, suites de narrations, etc.) ». En effet, les productions textuelles scolaires sont des *dérivés* des genres communs (narration, compte-rendu, etc.) mais transposés à l'école après avoir été « détachés » de leur sphère sociale de référence. Cette transposition scolaire entraîne des normes implicites dans l'élaboration des textes par les élèves : brièveté, obligation d'aboutissement final, etc. Ces contraintes « façonnent les textes scolaires et leur donnent un « air de famille » » (Boré & Elalouf, 2017 : 42). Mais cela ne suffit pas pour les considérer comme appartenant au genre de référence. À cela s'ajoute le fait que le texte narratif est le plus répandu à l'école, et de fait le plus représenté dans les corpus scolaires, notamment écologiques. Cette distribution inégale des situations de production alimente encore la question des genres. Comme le fait remarquer Roubaud (2017 : 116) : « il est intéressant de constater que même avec une longue expérience de constitution de corpus, les problèmes relatifs à la diversité des données demeurent […] ».

En outre, dans les textes d'élèves, plusieurs scripteurs partagent l'espace textuel : l'élève d'abord, en tant que scripteur originel, mais aussi l'enseignant lorsque celui-ci corrige ou annote la copie de l'élève. Ou encore, le scripteur lui-même lorsqu'il revient sur son texte dans une posture métatextuelle ou les scripteurs lorsque le texte est écrit en groupe. Ces

différentes interventions font des textes scolaires des genres dialogiques et multi-auctoriaux (Boré & Elalouf, 2017).

Ainsi, le caractère instable et hybride des textes composant les corpus d'écrits scolaires (Boré & Elalouf, 2017) participe à la difficile homogénéisation de ce type de corpus.

La deuxième difficulté soulevée par les corpus de textes scolaires est d'ordre méthodologique. En effet, comme le soulignent Doquet, David & Fleury (2017a : 8) :

> « […] malgré les avancées considérables des outils informatiques d'analyse de textes et les méthodologies liées aux grands corpus oraux, le traitement quantitatif de données langagières émanant de scripteurs débutants ou en cours d'apprentissage est difficile, en particulier du fait du caractère linguistiquement peu normé (ou autrement normé) de leurs productions (Lüdelin, 2008 ; Hirschmann & *al.*, 2013). »

Les textes d'élèves sont des productions non (ou peu) normées puisqu'elles émanent d'apprentis scripteurs. Un tel caractère singulier peut rendre difficile la confrontation de textes d'élèves avec d'autres corpus plus conformes à la norme, tels que les corpus de référence (Boré, 2007).

Afin de pouvoir réaliser une étude fine à la fois quantitative et qualitative d'un grand nombre de textes, un traitement automatique du corpus *via* des outils d'aide à l'annotation peut être réalisé, comme c'est le cas en linguistique de corpus. Les outils informatiques permettent d'exploiter de grandes masses de données en automatisant les calculs impossibles à la main (ou au mieux, complexes et très longs). Pour l'heure, dans la plupart des types de corpus, beaucoup de logiciels permettent de traiter des unités choisies en fonction de l'objectif de recherche (catégories morpho-syntaxiques, pragmatiques, sémantiques ou discursives), voire de mettre en relation plusieurs unités et visualiser les chaînes ainsi produites (De Vogüé & *al.*, 2017). Mais l'utilisation de ce type d'outil d'aide à l'annotation de corpus soulève deux difficultés lorsqu'il est question de les appliquer à des textes scolaires.

Premièrement, les outils d'aide à l'annotation existants nécessitent une version informatisée des textes à analyser. Or, le plus souvent, les textes scolaires sont manuscrits. Le passage d'un texte manuscrit à version informatisée nécessite donc un traitement spécifique (cf. ce chapitre section 5.2.2. et chapitre 6).

Deuxièmement, les outils d'aide à l'annotation nécessitent des textes non seulement informatisés, mais aussi normalisés du point de vue orthographique et dans une moindre

mesure, du point de vue de la ponctuation. Or, les textes d'élèves ne le sont pas, puisqu'il s'agit d'apprentis scripteurs. Ainsi, les écarts à la norme « constituent un obstacle à l'étiquetage automatique puisque les formes non normées ne sont pas reconnues par le [logiciel d'aide à l'annotation]» (David & Doquet, 2016 : 6). Cependant, produire une version normalisée des textes d'élèves suppose de rétablir orthographe et ponctuation. Comme le soulignent Boré & Elalouf (2017 : 44), le rétablissement de la norme fait « perd[re] ainsi les graphies et choix originaux du scripteur ». En outre, les écarts à la norme sont caractéristiques de la langue écrite des élèves. S'ils sont corrigés, ils influent sur l'authenticité des discours analysés et donc sur celle des corpus constitués (David & Doquet, 2016). Il faut alors trouver une solution pour que le logiciel d'annotation analyse la forme normée sans pour autant que la forme initialement produite par l'élève ne soit plus disponible (David & Doquet, 2016). Comme nous allons le voir, la meilleure solution consiste donc à garder une trace du texte originel *via* sa numérisation et de produire une ou plusieurs transcriptions que le logiciel d'annotation choisi pourra utiliser.

5.2.2. La transcription : une étape incontournable ?

La transcription de manuscrit est une étape essentielle par laquelle passent les philologues et les généticiens du texte pour analyser finement un texte littéraire dans son élaboration. Avec la constitution de grands corpus de textes scolaires, c'est-à-dire des textes manuscrits, cette étape devient nécessaire pour l'analyse linguistique, tant quantitative que qualitative. En effet, dès lors que l'on dispose de vastes corpus de textes d'élèves, la question de leur investigation se pose. Le chercheur peut être assisté dans cette tâche par des outils informatiques, mais cela passe par une étape préalable de numérisation et de transcription des textes car les écrits non informatisés empêchent toutes sortes de requêtes : l'investigation sur les textes reste limitée, d'autant plus lorsqu'il s'agit d'un grand nombre de textes (Roubaud, 2017). La transcription de ces textes devient alors nécessaire. Pour présenter les enjeux de la transcription de texte, nous nous appuierons principalement sur la synthèse de Testenoire (2017), dont les travaux sur les brouillons d'écrivains peuvent être rapprochés des textes d'apprentis scripteurs, bien que la première à avoir fait le pari d'analyser des textes d'élèves avec les mêmes méthodes que les textes d'écrivains est C. Fabre-Cols.

Selon Testenoire (2017), la transcription de manuscrits vise deux objectifs : restituer la disposition et les caractéristiques visuelles de l'écrit manuscrit (soulignements, ratures, disposition spatiale de l'écrit…) et présenter au lecteur (ou en l'occurrence, au logiciel d'annotation), le texte le plus clair et accessible possible, c'est-à-dire conforme aux conventions typographiques usuelles. Ainsi, « les choix de transcription sont une négociation entre ces deux exigences contradictoires » (Testenoire, 2017 : 91).

En effet, les choix de transcription dépendent de plusieurs facteurs. D'abord, l'objectif de la transcription détermine le choix des paramètres de l'écrit qui seront retenus. Par exemple, dans la constitution du corpus *Ecriscol*, le projet d'étudier les erreurs d'orthographes dans les écrits d'élèves a mené les chercheurs à faire le choix d'une transcription reproduisant l'orthographe non normée des textes scolaires.

Le deuxième facteur qui influe sur les choix de transcription est l'objet transcrit. Les manuscrits permettent une liberté de formes et de contraintes que l'écrit informatisé restreint : graphies différentes et possibilités de dessins ou signes, utilisation non linéaire de l'espace de la feuille, affranchissement de normes linguistiques et typographiques….

Enfin, l'outil employé pour la transcription la contraint nécessairement, selon qu'il donne la possibilité ou non d'intégrer des images, des ressources hypertextuelles, etc.

Ces variables déterminent le type de transcription à adopter. La génétique textuelle, dont l'objectif est la reconstitution du processus d'écriture, distingue trois types de transcriptions : la transcription diplomatique, linéarisée et chronologique.

La transcription diplomatique « vise la reproduction exacte du manuscrit » (Testenoire, 2017 : 94). Elle est donc la moins interprétative des trois puisqu'elle ressemble à une photographie de l'original (Doquet & *al,* 2017b).

La transcription linéarisée « consiste à transcrire tous les signifiants graphiques présents sur le manuscrit dans une ligne continue de signes. Ce qui est hors ligne est linéarisé à l'aide d'un codage typographique. » (Testenoire, 2017 : 94). Le codage typographique s'appuie sur divers types de conventions[48].

[48] Les différents types de conventions peuvent être, par exemple :

- les conventions utilisées par l'Institut des Textes et Manuscrits (ITEM) (David & Doquet, 2016 ; Doquet & *al.*, 2017b ; Testenoire, 2017) : les chevrons < > pour signifier un segment de texte ajouté ou les crochets *[ ]* pour une suppression

La transcription linéarisée repose donc sur l'interprétation du transcripteur puisqu'elle remet sur un axe linéaire la succession des opérations d'écriture (Doquet & *al,* 2017b).

Enfin, la transcription chronologique : « repose sur le principe de la substitution génétique auxquelles sont ramenées les quatre opérations d'écriture fondamentales (ajout, suppression, remplacement et déplacement). Elle consiste donc à reconstituer les différents états de l'écrit entre lesquels est intervenue une substitution génétique. » (Testenoire, 2017 : 94). Ce dernier type de transcription privilégie la restitution des données temporelles et met au jour les différentes strates de l'écriture (Doquet & *al.,* 2017b).

Comme le soulignent Doquet & *al.* (2017b), les principes de la transcription ainsi posés ne la rendent pas plus aisée pour autant, d'autant plus lorsqu'il s'agit de la transcription des écrits d'élèves. Dans les faits, la transcription des textes d'élèves nécessite plusieurs versions. Une version scannée du texte, parce qu'elle le photographie tel qu'il est, peut être considérée comme une version iconique du texte. Cette étape est d'ailleurs attestée dans la plupart des corpus de textes scolaires (Elalouf & Boré, 2007, Auriac-Slusarczyk & *al.*, 2008, avec le corpus *Grenouille*, Doquet & *al.,* 2017, avec *Ecriscol*, Wolfarth & *al.,* 2017, avec *Scoledit*…).

Une transcription diplomatique mêlant des principes de transcription linéarisée rend témoignage des modifications en cours de rédaction ou lors de la relecture grâce à des signes et conventions tels que ceux décrits *supra*. Cette version est à privilégier pour les écrits d'élèves qui ne comportent pas d'intervention de l'enseignant (Elalouf, 2004 ; Elalouf & Boré, 2007).

Pour les textes comportant à la fois les interventions de l'élève et celles de l'enseignant, une version commentée est préférable car elle permet de voir comment l'élève interprète les annotations du professeur. Elle se traduit par un autre système de signes et conventions qui distinguent les interventions de l'élève de celles du professeur, par exemple en codant *P* les commentaires du professeur puis en les annotant dans une police différente. Si ce type de transcription a l'avantage de tout expliciter, elle est peu naturelle et ralentit considérablement la lecture du texte (Elalouf, 2004 ; Elalouf & Boré, 2007).

- les conventions utilisées dans la norme internationale TEI-P5 (Garcia-Debanc & *al.*, 2017 ; De Vogüé & *al.*, 2017)
- les conventions adaptées des conventions de transcription de l'oral (Roubaud, 2017).

Enfin, une version orthographiée selon la norme et éventuellement dépouillée des dessins, tableaux et autres signes difficilement reproductibles informatiquement, peut également intervenir, soit pour rendre plus clairs les textes nécessitant un effort de lecture important et rendre plus visibles les phénomènes propres au processus d'écriture, soit pour soumettre le texte à un logiciel d'aide à l'annotation. La version orthographiée selon la norme contraint le transcripteur à des choix, comme le mettent en évidence Boré & Elalouf (2017 : 38) :

> « où s'arrêter dans la réécriture quand interfèrent des choix graphiques, morphosyntaxiques, avec leur contrepartie sémantique ? Est-on autorisé à remplacer *il rentrer chez eux chacun* par *ils rentrèrent chez eux chacun* ; *il se partagea la somme* par *ils se partagèrent la somme* ? Et faut-il restituer une ponctualité forte avant ? »

La transcription d'un corpus de textes scolaires peut ainsi être composée de plusieurs versions transcrites d'un même texte, comme autant de strates de transcription servant à l'annotation du ou des phénomène(s) que l'on cherche à observer. Cependant, dans le cas de grands corpus de textes scolaires dont une des finalités est le partage avec la communauté scientifique, il est possible d'envisager une transcription hétérogène qui prendrait en compte le plus d'éléments possible, mêlant ainsi les divers types de transcriptions (diplomatique, linéarisée, chronologique). C'est le cas par exemple du corpus *Ecriscol*. Dans ce projet, les conventions retenues transcrivent à la fois des caractéristiques sémiotiques et topographiques (ajouts, suppressions, ratures, changement de graphie, localisation hors ligne d'un élément, etc.), mais aussi chronologiques (codage pour différents temps dans l'écriture) et ainsi que les commentaires de l'enseignant (Testenoire, 2017).

Ainsi, la transcription de textes d'élèves relève d'une grande variété de facteurs, comme le souligne Testenoire (2017 : 92) :

> « Une telle diversité explique en partie l'absence d'harmonisation des conventions de transcription, régulièrement déplorée. La nécessité d'adapter les conventions de transcription aux problèmes posés par chaque corpus manuscrit prime bien souvent dans les faits sur le souhait d'harmonisation. »

L'absence de convention de transcription universelle est un élément supplémentaire qui engendre des difficultés lors de la constitution et de l'analyse des corpus d'écrits scolaires.

5.3. L'exploitation de corpus

L'analyse d'un corpus qui serait transcrit sans faire l'objet d'une annotation fine, au moyen d'un outil d'aide à l'annotation ou non, perdrait de sa richesse. En effet, l'annotation consiste à ajouter de l'information aux données langagières (Habert, 2005 ; Doquet & *al.*, 2017b). Il s'agit donc d'un acte interprétatif d'intérêt linguistique dont l'objectif est un corpus enrichi ou simplifié (Leech, 1997). Plusieurs lectures d'un même texte peuvent être envisagées en fonction de l'angle d'étude choisi : syntaxique, sémantique, etc. Les différentes lectures, traduites dans les faits par un étiquetage spécifique des entités choisies, constituent autant de strates d'annotation (Fort, 2012). On peut d'ailleurs se demander, à l'instar de Doquet & *al.* (2017b), dans quelle mesure ce que nous appelons *transcription* n'est pas déjà une annotation, étant donné qu'une transcription est le résultat de choix du transcripteur et l'ajout d'informations à un texte originel.

En 5.3.1., nous décrivons succinctement quelques outils d'aide à l'annotation utilisés pour l'exploitation de corpus de textes scolaires. En 5.3.2., nous passons en revue de manière non exhaustive les différents corpus constitués en vue d'étudier la référence.

5.3.1. Quelques outils d'aide à l'annotation

Pour assister la tâche d'exploitation d'un corpus composé d'un nombre important de textes, il est possible d'utiliser un outil d'aide à l'annotation. Il s'agit d'une « interface facilitant l'annotation manuelle des données » (Fort, 2012 : 33) en proposant une lecture des textes annotés et des fonctionnalités précises (en plus des calculs statistiques communs à tous les logiciels) comme par exemple l'accord inter-annotateur lorsque plusieurs personnes annotent le même corpus. Chaque outil possédant ses spécificités, le logiciel choisi sera fonction du phénomène linguistique étudié et du point de vue de l'analyse (syntaxique, morpho-syntaxique, sémantique, etc.).

Dans l'annotation de corpus de textes scolaires, plusieurs outils ont été utilisés sur un même matériau. Il peut s'agir de logiciels ordinaires tels qu'un traitement de texte pour effectuer des recherches lexicales dans le texte (Elalouf & Boré, 2007), d'un vérificateur d'orthographe pour détecter des erreurs ou d'un tableur pour des comptages statistiques

communs comme le nombre d'apparition d'une occurrence. Pour une analyse plus fine, les outils communs précédemment cités peuvent être combinés à l'utilisation de logiciels plus spécifiques.

Dans les travaux d'Elalouf & Boré (2007), dont l'objectif est de mieux caractériser l'écriture scolaire, le logiciel d'aide à l'annotation *Tropes* a ainsi été retenu car « il effectue, au terme de l'analyse automatique, un diagnostic du style général du texte et de sa mise en scène, en fonction des indicateurs statistiques collectés » (Elalouf & Boré, 2007 : 60). Dans ces travaux, *Tropes* s'appuie sur la catégorie sémantique des verbes pour déterminer le type de mise en scène de l'histoire racontée par l'élève. Par exemple, si les verbes d'action sont fréquents dans le texte, la mise en scène sera qualifiée de dynamique par le logiciel. *Tropes* s'appuie également sur la fréquence des pronoms de première personne pour aider à conclure sur le mode d'énonciation choisi par l'élève.

Dans le projet *Ecriscol* (Doquet & *al.*, 2017b), a été utilisé le logiciel de textométrie *Le Trameur*. Initialement conçu pour compter des unités dans les textes ou dans les différentes parties d'un texte afin de mesurer et de contraster leurs comportements, le logiciel peut également « donner à voir [les] différentes « couches » textuelles dans ces différents modules d'édition du corpus » (Doquet & *al.*, 2017b : 150). Il permet ainsi d'examiner chaque niveau d'annotation et les opérations de calcul correspondantes.

Plusieurs autres outils d'aide à l'annotation existent, qu'ils soient plutôt généraux ou présentent des spécificités. Le logiciel *Glozz* (Widlöcher & Mathet, 2009) par exemple, permet d'explorer et d'annoter un corpus au niveau discursif (délimitation de segments textuels, identification de connecteurs de discours, etc.). Le logiciel *Analec,* (Landragin & *al.* 2012) permet de visualiser et d'étudier plus particulièrement les chaînes de référence. Plusieurs outils peuvent être combinés, dans l'objectif d'affiner toujours plus l'analyse du corpus. Cependant, à notre connaissance, aucun d'entre eux n'a encore été mis à l'épreuve d'un corpus de textes d'élèves.

5.3.2. Les corpus constitués en vue d'étudier la référence: un manque de corpus d'écrits scolaires

Lorsque l'on se place du point de vue qui nous intéresse particulièrement dans ce travail de thèse, à savoir l'étude de la référence et plus spécifiquement, des anaphores et des chaînes de référence (CDR) dans les écrits d'élèves, nous constatons qu'il n'existe pas à l'heure actuelle de corpus de textes d'élèves constitué dans le but d'étudier ces phénomènes.

Avant d'aller plus loin, il nous faut distinguer deux types d'études sur la référence. D'une part, les travaux portant sur l'analyse linguistique des expressions référentielles, tels que ceux décrits dans les chapitres de notre cadre théorique. D'autre part, les travaux portant sur l'élaboration de ressources dont l'objectif est l'étude des phénomènes référentiels.

En effet, un certain nombre de travaux en TAL ont pour objectif la constitution de corpus d'étude de la référence afin d'entraîner des modèles informatiques à la reconnaissance des phénomènes d'anaphore, de coréférence, etc. Ces modèles informatiques, une fois entraînés, servent de base de fonctionnement à des outils d'aide à l'annotation dont la spécificité est l'étude de la référence. Dans les corpus recueillis à des fins d'entraînement de modèles informatiques, on peut citer pour l'anglais par exemple, les corpus MUC-6 et MUC-7 (*Message Understanding Conference*) (Chinchor & *al.*, 1997). Ces deux corpus sont annotés en vue du traitement de la coréférence entre noms, groupes nominaux et pronoms (personnels et démonstratifs). On peut citer également le corpus annoté par Poesio & Vieira (1998), issu du *PennTree-bank*[49]. Le corpus de Poesio & Vieira regroupe plusieurs articles de presse dans lesquels sont annotés environ 1400 descriptions définies. Ces annotations ont permis de mieux quantifier les différentes catégories de descriptions définies et d'identifier quelques difficultés dans l'annotation et la résolution des descriptions définies (Gardent & Manuelian, 2005). Cependant, ces travaux concluent « qu'il est difficile voire impossible d'annoter les descriptions définies d'un corpus de façon suffisamment fiable pour que ce corpus puisse servir à l'évaluation et à l'entraînement de modules d'interprétation des expressions référentielles » (*ibid.* : 3).

[49] Le *Penn Treebank* a été créé par l'Université de Pennsylvanie entre 1989 et 1992. Il a d'abord été question de la partie morpho-syntaxique (*part-of-speech tagging),* qui comprend trente-six catégories principales et douze pour la ponctuation et les symboles. Puis en 1994, il a été augmenté d'une partie sur la syntaxe (*bracketing*) qui comprend une quinzaine de catégories.

Les corpus de langue française qui permettent d'étudier les expressions référentielles sont peu nombreux et concernent tous des textes d'adultes[50]. De surcroît, aucun d'entre eux n'est suffisamment important pour faire référence. Nous pouvons tout de même citer le corpus ARCADE (Tuttin & *al.*,2000 a et b) et son million de mots, dans lequel toutes les expressions anaphoriques d'articles (de presse ou de recherche) sont annotées (pronoms personnels, déterminants possessifs, démonstratifs, indéfinis…) excepté les descriptions définies. L'étude montre que les procédés anaphoriques varient d'un type de texte à l'autre même dans un ensemble de textes appartenant *a priori* au même genre textuel. On retrouve l'étude particulière des descriptions définies dans le corpus *DEDE,* contenant environ 48 000 mots (Gardent & Manuelian, 2005). L'objectif visé à travers l'annotation de ce corpus est l'interprétation automatique des descriptions définies, c'est-à-dire les expressions de la forme *le/la/les N.*

Plus précisément, lorsqu'il s'agit de corpus permettant l'étude des chaînes de référence, Schnedecker & *al.* (2017 : 9) font le constat suivant :

> « […] il n'existe pas, pour le français, de corpus longitudinal annoté où soit prise en considération l'évolution de la composition des CR, alors qu'il existe des corpus annotés morpho-syntaxiquement de textes issus de toutes les périodes de la langue française. »

Les chercheurs parlent ici d'un corpus diachronique qui s'étendrait sur plusieurs époques du français écrit (de l'ancien français au français contemporain par exemple). Cependant, le constat est le même d'un point de vue développemental : il n'existe pas à l'heure actuelle de corpus annoté qui étudie l'évolution de la composition des CDR en fonction de l'âge des enfants et de l'évolution de leurs compétences discursives en matière de cohésion textuelle.

Dans ce contexte, un projet du laboratoire CLLE-ERSS[51] dans le cadre de l'ANR E:CALM[52], et dont notre travail de thèse est concomitant, est en train de voir le jour. Suite aux constats de la rareté des grands corpus d'écrits scolaires et de l'absence de corpus d'apprenants dédiés à l'analyse de la cohésion textuelle (dont la référence est une manifestation, cf. chapitres 1 et 2), le corpus RESOLCO (Résolution de problèmes de cohésion textuelle) vise plusieurs objectifs. Le premier objectif est de structurer et mettre à la

[50] Nous renvoyons à toutes les études détaillées dans le chapitre 2 (notamment Condamines, 2005 ; Longo & Todirascu, 2009 ; Schnedecker & Longo, 2012 ; Longo, 2013 ; Landragin & *al.* 2015 ; Charolles & Storme, 2015).

[51] CLLE-ERSS – Cognition, Langues, Langage, Ergonomie – Université Toulouse 2.

[52] E :CALM : Ecriture scolaire et universitaire : Corpus, Analyses Linguistiques, Modélisation didactiques.

disposition de la communauté scientifique un vaste corpus d'écrits d'élèves et d'étudiants permettant des analyses quantitatives et des traitements outillés. Le deuxième objectif est de produire des outils pour le traitement de ces corpus (guides de transcription, guides d'annotation). Enfin, il s'agit aussi de caractériser ces écrits du point de vue de l'acquisition de la cohésion textuelle (annotation discursive) aux différents niveaux de cursus et déterminer des indicateurs de progressivité pour la cohésion textuelle. Le corpus RESOLCO recueille à ce titre environ 500 textes d'élèves de l'école primaire au collège (du CE2 à la 3°) et des textes d'étudiants. Les textes sont complétés par des annotations d'enseignants et des réécritures. Tous les textes répondent à la même consigne : les élèves doivent raconter une histoire dans laquelle ils insèrent trois phrases qui leur sont imposées. Nous reviendrons en détail sur cette consigne, qui est aussi celle de notre travail de thèse, dans les chapitres 6 et 7.

SYNTHESE des notions centrales du chapitre 5

Dans ce chapitre, nous avons voulu montrer les enjeux de la constitution de grands corpus d'écrits scolaires ainsi que les difficultés inhérentes à ce type de matériau.

Dans la première partie de ce chapitre, nous avons montré que si de nombreux corpus de référence existent en linguistique de corpus, les grands corpus qui étudient les écrits d'apprenants sont nettement moins nombreux et mieux représentés en français langue étrangère qu'en français langue première (Granger, 2009 ; Bilger & Cappeau, 2013 ; Wolfarth & *al.*, 2017). La rareté attestée des grands corpus d'écrits d'apprenants tient au fait que pendant longtemps, les chercheurs ont travaillé sur un échantillon de textes limité (David, 2000a ; Elalouf, 2004 ; Elalouf & Boré, 2007 ; Garcia-Debanc & *al.,* 2017 ; Roubaud, 2017) et analysé un phénomène linguistique donné en fonction de leur objectif d'étude (Gunnarsson-Largy & Auriac-Slusarczyk, 2013). La prise de conscience de la richesse d'une exploitation qui compare divers angles d'études et qui s'appuie sur un nombre représentatif de textes d'élèves (Auriac-Slusarczyk & *al.*, 2008 ; Elalouf, 2011 ; Masseron, 2011 ; Doquet & *al.*, 2017a) a conduit à plusieurs tentatives de constitution de grands corpus de textes scolaires.

En effet, la constitution d'un grand corpus d'écrits scolaires a pour objectif de mener une analyse à la fois linguistique et didactique des productions recueillies, et de constituer des ressources utilisables par les chercheurs mais aussi par les formateurs d'enseignants et les

enseignants eux-mêmes (Elalouf, 2011 ; Garcia-Debanc & *al.*, 2017 ; Jacques & Rinck, 2017). Les travaux précurseurs d'Elalouf & Boré (Elalouf & *al.*, 2005 ; Elalouf & Boré, 2007) ont été suivis par d'autres projets qui s'appuient sur des corpus toujours plus importants : *corpus Grenouille* (Auriac-Slusarczyk & *al.*, 2008 ; Gunnarsson-Largy, & Auriac-Slusarczyk, 2013), projet *Ecriscol* (David & Doquet, 2016 ; Doquet & *al.*, 2017b), projet *Scoledit* (Wolfarth & *al.*, 2017), corpus de littératie avancée (Jacques & Rinck, 2017). La visée de tels corpus est de « réunir des corpus représentatifs, comparables, cumulables » car c'est « un enjeu essentiel pour la production de connaissances en didactique du français. » (Elalouf, 2011 : 57).

Toutefois, réunir un tel matériau s'avère complexe, comme nous l'avons exposé dans la deuxième partie de ce chapitre. En effet, les textes scolaires comportent des spécificités et suscitent des difficultés de plusieurs ordres.

D'abord, un *corpus* est constitué de textes présentant une certaine unité de genre. Or, les productions textuelles scolaires sont des *dérivés* des genres communs (narration, compte-rendu, etc.) mais transposés à l'école après avoir été détachés de leur sphère sociale de référence (Boré, 2007 ; Boré & Elalouf, 2017).

Une difficulté d'ordre méthodologique apparaît ensuite : les textes d'apprenants ne respectent pas la norme orthographique ni ponctuationnelle (Lüdelin, 2008 ; Hirschmann & *al.*, 2013). Or pour être analysés, ils doivent être soumis à un traitement quantitatif *via* des outils informatiques d'analyse de texte qui ne savent traiter que des textes normés (David & Doquet, 2016 ; De Vogüé & *al.*, 2017). Dès lors, le passage par une ou plusieurs transcriptions est nécessaire.

La question de la transcription engendre également quelques difficultés étant donné les nombreux facteurs qui entrent en jeu dans la transcription de textes : l'objectif de la transcription, l'objet transcrit, l'outil utilisé pour transcrire mais aussi le type de transcription souhaitée (diplomatique, linéarisée, chronologique) (Fabre-Cols, 2000 ; Testenoire, 2017). La recension de plusieurs méthodologies de transcription des divers corpus de textes scolaires précédemment cités a montré le manque d'harmonisation des pratiques en la matière, au travers notamment de la grande variété de conventions et des principes de transcription utilisés (version iconique, transcription diplomatique, transcription linéarisée…) (Doquet & *al.*, 2017 ; Garcia-Debanc & *al.*, 2017 ; Roubaud, 2017).

Malgré l'hétérogénéité des pratiques, toutes les transcriptions sont soumises à la même dernière étape du processus d'exploitation de corpus. Il s'agit de l'annotation. Cette étape aussi peut être source de difficultés, notamment dans le choix des outils d'aide à l'annotation. Bien que l'utilisation d'un logiciel commun tel que traitement de texte, tableur ou vérificateur d'orthographe soit pertinente, il est judicieux de compléter l'analyse par des outils spécifiques (Elalouf & Boré, 2007 ; Landragin, 2014). L'angle d'étude de chaque corpus déterminera l'outillage informatique : le logiciel *Tropes* pour le corpus d'Elalouf & Boré (2007), *Le Trameur* dans le projet *Ecriscol* ou d'autres encore tels *Glozz* ou *Analec* dans d'autres types de corpus.

Nous avons terminé ce chapitre par le constat qu'il n'existe pas de corpus annoté pour l'étude de la cohésion textuelle, et plus particulièrement des expressions référentielles et des chaînes de référence et (Schnedecker & *al.*, 2017), sur des textes d'apprenants. Dans ce contexte, nous avons présenté le projet RESOLCO en cours et concomitant de notre travail de thèse, dont l'objectif est l'étude de l'acquisition de la cohésion textuelle à différents niveaux de la scolarité.

CHAPITRE 6 : Le recueil de données et la constitution du corpus d'étude

Afin d'étudier l'acquisition des marques de la référence chez les élèves de plusieurs niveaux, nous avons déterminé des questions de recherche et constitué un corpus d'étude. En 6.1., nous présentons donc la problématique et nos questions de recherche. En 6.2., nous exposons les données recueillies et rendons compte des moyens mis en œuvre pour leur collecte.

Une fois les données collectées, il est nécessaire de les transcrire pour qu'elles puissent devenir un objet d'analyse et constituer un corpus d'étude partageable et analysable *via* des outils d'annotation. En 6.3., nous abordons les difficultés soulevées et les choix méthodologiques mis en œuvre, tant pour les éléments qui relèvent de la transcription diplomatique que de la transcription normée.

6.1. Problématique et questions de recherche

Dans la première partie de notre étude, nous avons posé le cadre théorique dans lequel s'inscrit notre recherche, sur les plans linguistiques, psycholinguistiques et didactiques. Parmi les différentes composantes qui constituent l'activité complexe de la construction du discours chez les élèves, nous avons fait le choix de nous centrer sur la cohésion textuelle, et particulièrement sur les marques de la continuité référentielle. Plus précisément, en tant que marque de continuité référentielle, nous nous intéressons à l'anaphore et aux chaînes de référence.

Les phénomènes d'anaphore ont été abondamment étudiés du point de vue de l'analyse et de la réception chez le locuteur adulte. En témoignent des travaux comme ceux de Charolles (1978), Reichler-Béguelin (1988a), Kleiber (1990), Apothéloz (1995a), Corblin (1995), ou encore Cornish (2009). Ces études ont apporté de nombreux éléments de compréhension et de description du fonctionnement des anaphores. Cependant, un nombre moins grand de recherches se sont intéressées à l'étude de la référence du point de vue de la production, et particulièrement chez les enfants (Charolles, 1988c ; Schnedecker, 1995 ;

Masseron & Schnedecker, 1995). Ces travaux se sont principalement attachés à décrire les modalités de l'expression référentielle dans les textes d'élèves, notamment le choix des marques anaphoriques pour introduire et maintenir les référents. Les riches données qualitatives apportées concernent toutefois un moment donné de la scolarité, dans une approche plutôt « instantanée » (Schnedecker, 1995) et non dans une démarche de progressivité qui analyse une évolution sur plusieurs niveaux.

Le caractère de progressivité est central dans la problématique de la construction du discours chez les élèves. En effet, définir les compétences dont l'acquisition est possible pour les élèves d'un niveau scolaire donné permet d'établir une cartographie des connaissances et des compétences tout au long du cursus scolaire. Toutefois, en didactique du français langue première, peu de travaux se placent dans cette perspective (Nonnon, 2010). Ce constat peut s'expliquer en partie par la rareté du matériau exigé pour conduire ce type d'étude. Étudier l'acquisition de compétences dans une perspective de progressivité nécessite d'analyser un grand nombre de données. Or, les « grands » corpus de textes scolaires sont très peu nombreux. En didactique du français langue première, les textes scolaires font généralement l'objet d'une analyse fine sur un petit échantillon de textes (Fabre-Cols, 2000 ; Masseron, 2005 ; Boré, 2011).

On trouve toutefois le caractère de progressivité dans les travaux psycholinguistiques sur le développement de l'organisation discursive des enfants. Plus particulièrement, plusieurs recherches montrent une évolution progressive dans la maîtrise des marques de la référence à l'oral comme à l'écrit (Karmiloff-Smith 1979, 1981 notamment). L'intérêt d'une étude développementale est donc de savoir quelles marques anaphoriques les enfants utilisent pour introduire et maintenir les référents et à quel moment ces marques apparaissent dans les différentes phases du développement (à quels âges).

Cet ensemble de constats nous a conduite à définir les contours de notre recherche qui a pour objectif l'étude et la description de l'acquisition des marques de cohésion textuelle, et plus précisément, des marques de la référence, dans des productions écrites d'élèves du CE2 au CM2, soit de 9 à 11 ans.

Notre question de recherche principale est donc la suivante : **comment les élèves du CE2 au CM2 introduisent-ils et maintiennent-ils les référents dans des productions écrites narratives ? Y a-t-il des différences selon les niveaux de classe ?**

Les réponses que nous apporterons seront susceptibles de fournir :

- Sur les plans linguistiques et psycholinguistiques : une cartographie des marques de la référence privilégiées par les élèves, en fonction de leur âge
- Sur le plan didactique : une meilleure compréhension du schéma de maîtrise de la référence par les élèves afin d'une part, de servir d'appui au jugement de cohérence des enseignants ainsi qu'à l'évaluation, et d'autre part, d'identifier des points d'entrée pour l'intervention didactique

Notre question générale appelle d'autres questionnements, auxquels la présente étude se propose de répondre.

Au niveau de l'introduction des référents, on cherche à déterminer les différents degrés de maîtrise dans la manière d'introduire des référents, en fonction du niveau de classe des élèves : quel type d'expression référentielle emploient-ils en première mention ? Le caractère humain/non humain ou animé/non animé des référents a-t-il une influence sur le choix des formes linguistiques de la première mention ? Les élèves introduisent-ils les référents de manière locale ou bien globale ? Nous faisons l'hypothèse que les formes linguistiques employées pour introduire un référent dépendent de la nature de l'entité en question et, dans une certaine mesure, du niveau de classe.

Au niveau du maintien de la référence, on cherche à déterminer les différentes stratégies employées par les élèves dans les chaînes de référence en fonction du niveau de classe : pour référer, les élèves utilisent-ils des anaphores (2 mentions) ou bien de véritables chaînes de référence (3 mentions et +) ? Quelle est la composition des chaînes de référence (longueur moyenne, type d'expressions référentielles) ? La nature des référents influe-t-elle sur la longueur de la chaîne de référence ? Nous faisons l'hypothèse que le nombre d'expressions référentielles employées pour maintenir la référence (anaphore ou CDR) dépend du niveau de classe et que la nature des référents influe sur la longueur et la composition de la chaîne.

Lorsqu'il y a des ambiguïtés, on cherche à comprendre en quoi le non-respect des règles (cognitives et linguistiques) de la cohérence participe à l'ambiguïté référentielle d'un texte.

L'ensemble de ces éléments constituent des indices utiles pour répondre à notre question de recherche. En tant que tels, ils doivent être considérés comme des critères d'analyse pertinents pour notre étude, sur lesquels nous reviendrons plus en détail dans le chapitre 7.

6.2. Recueil des données

Une part importante de la constitution d'un corpus d'étude est le recueil de données. Nous présentons en 6.2.1. notre protocole expérimental. Les raisons qui ont déterminé le choix de notre public et les conditions de recueil des données sont exposées en 6.2.2. Nous terminons cette section par la présentation de la composition de notre corpus d'étude final.

6.2.1. Protocole expérimental

Pour mesurer les marqueurs linguistiques utilisés par les élèves pour désigner un même référent, nous reprenons une tâche problème de C. Garcia-Debanc contraignant les élèves à résoudre plusieurs problèmes d'anaphores (Garcia-Debanc, Bonnemaison, 2014 ; Garcia-Debanc, Bras, 2016 ; Garcia-Debanc & *al.* 2017). La consigne d'écriture s'inspire de celle de Charolles (1988 ; cf. chapitre 3, section 3.1.2.3.) mais se distingue par le support inducteur choisi. Là où Charolles (*ibid.*), dans la lignée des travaux en psycholinguistique, s'appuie sur un ensemble d'images composant une bande dessinée, notre consigne s'appuie sur un ensemble de phrases. Ce choix nous permet d'analyser la prise en compte de certains indices linguistiques par les élèves dans la résolution de la tâche. De ce fait, elle peut être qualifiée d'épilinguistique car elle attire l'attention des élèves sur une dimension du fonctionnement de la langue, sans pour autant utiliser les termes du langage technique (Garcia-Debanc & *al.* 2017).

La consigne de la tâche-problème proposée est la suivante :

> *Racontez une histoire dans laquelle vous insèrerez, séparément et dans l'ordre donné, les trois phrases suivantes :*
>
> *1. Elle habitait dans cette maison depuis longtemps.*
>
> *2. Il se retourna en entendant ce grand bruit.*

3. Depuis cette aventure, les enfants ne sortent plus la nuit.

La tâche-problème comporte plusieurs problèmes à résoudre. Pour en proposer l'analyse, nous reprenons Garcia-Debanc & Bonnemaison (2014) et Garcia-Debanc & *al.* (2017).

Les phrases de la consigne contiennent des expressions référentielles anaphoriques qui nécessitent l'introduction de référents : les pronoms personnels *elle* et *il*, les syntagmes nominaux (SN) démonstratifs anaphoriques *cette maison* et *ce grand bruit*, le SN défini pluriel *les enfants*. Ainsi, l'absence de référence explicite pour chaque occurrence à traiter conduit les élèves à insérer des phrases supplémentaires avant l'occurrence. Toutefois, les distances entre une reprise anaphorique et un terme source peuvent être diverses. La résolution peut être effectuée selon des enchaînements locaux entre phrases successives ou nécessiter une plus grande distance entre ces deux éléments.

Si la tâche proposée induit les élèves à gérer les marques de cohésion textuelle, elle ne donne aucune indication sur les personnages. Les élèves doivent décider eux-mêmes si les enfants mentionnés dans la dernière phrase sont un terme générique reprenant les personnages désignés précédemment par les pronoms personnels *elle* et *il,* s'il s'agit d'autres personnages ou encore, si cette moralité est une généralisation à partir de l'aventure arrivée à un seul enfant.

En outre, le choix du genre du récit est tout à fait ouvert. Il est guidé par la phrase fermoir, avec les mots *nuit* et *aventure* et le caractère de moralité qu'elle semble revêtir, qui suggère un danger et une peur. Comme l'indiquent Garcia-Debanc & *al.* (2017) : « ces éléments constituent donc à la fois des contraintes linguistiques à prendre en compte et des lanceurs d'écriture, tremplins de l'imaginaire ». Ainsi la dernière phrase joue-t-elle un rôle important pour la planification textuelle (Hayes & Flower, 1980), dans la mesure où elle contraint le cadre temporel de l'épisode-clé du récit, la nuit, et oriente l'atmosphère générale : si les enfants ne sortent plus la nuit, c'est qu'un ou plusieurs d'entre eux ont vécu une ou plusieurs aventures terrifiantes ou dangereuses la nuit. Aussi, tous les éléments induits par la phrase finale font apparaître la nécessité d'une planification à rebours du texte (Schneuwly, 1988).

Les deux premières phrases imposées de la consigne sont écrites à l'imparfait ou au passé simple, ce qui suppose que le texte à rédiger doit l'être dans les temps du passé. La troisième phrase imposée, quant à elle, est écrite au présent et a valeur « d'affirmation générique » (Garcia-Debanc & Bonnemaison, 2014). Elle implique de ce fait un lien de cause/conséquence entre le récit au passé comportant les deux premières phrases imposées et elle-même. De surcroît, l'élève doit aussi prendre garde à respecter les indications spatiales (*dans cette maison*) et temporelles *(depuis longtemps, la nuit, depuis cette aventure).*

Pour contourner d'éventuelles modifications apportées aux phrases proposées, celles-ci sont inscrites sur des bandelettes de papier que les élèves collent à l'endroit où ils le souhaitent dans le texte. Ce dispositif a été également choisi pour éviter d'induire la rédaction d'une longueur donnée de texte si ces phrases avaient été séparées par des pointillés.

En ce qui concerne le mode de passation, l'enseignant.e/l'expérimentateur a pour consigne d'expliquer le moins possible l'exercice et de s'en tenir strictement à la consigne distribuée, afin de ne pas biaiser les résultats en donnant des exemples pour aider les élèves ou en identifiant à leur place les occurrences à introduire. Chaque séance d'écriture a duré entre 40 et 50 minutes. Les élèves pouvaient recourir au brouillon s'ils en éprouvaient le besoin mais celui-ci n'était pas examiné par un adulte. De fait, la grande majorité des élèves a écrit directement le texte.

Dès la fin de la collecte, chaque texte a été scanné afin de conserver une iconicité maximale (Boré & Elalouf, 2017 ; cf. chapitre 5 section 5.2.2.) permettant de restituer le graphisme, la mise en page, les éventuels jeux de couleur. En outre, le support numérique permet d'accéder à la photographie de l'original en cas de doute ou pour une analyse plus fine du texte.

6.2.2. Public choisi et conditions de recueil

L'étude de la référence chez des élèves de 9 à 11 ans est la continuité de notre recherche de master. La thèse devait initialement porter sur l'étude de la référence chez les élèves de 9 à 15 ans, soit du CE2 à la troisième. Devant l'étendue du public concerné, nous avons restreint notre étude à l'observation de textes d'élèves de 9 à 11 ans, soit du CE2 au CM2. À l'époque de ce choix, l'Éducation Nationale groupait ces trois niveaux scolaires dans

le cycle 3 de l'école primaire. Depuis, le CE2 a été intégré au cycle 2 (CP, CE1, CE2) et le cycle 3 regroupe maintenant les CM1, CM2 et 6°. Malgré cette réorganisation officielle, nous avons fait le choix de garder notre public initial (CE2, CM1, CM2) car il permet de suivre de plus près les évolutions dans les apprentissages telles que celles constatées en psycholinguistique développementale entre 8/9 ans et 11 ans.

Le public visé par notre recherche se compose donc d'élèves de 9 à 11 ans, du CE2 au CM2 de l'école primaire française. Les élèves sont issus d'écoles primaires de Haute-Garonne et du Tarn-et-Garonne. Certaines écoles sont en milieu rural, d'autres en milieu urbain. Une école propose des enseignements bilingues français-occitan (Calandreta). En outre, certaines classes sont composées d'un petit nombre d'élèves et d'autres sont des classes multiniveaux. Il y a donc une hétérogénéité certaine du point de vue sociologique.

Le corpus ayant fait l'objet d'une extension progressive, la collecte des données a été effectuée en plusieurs temps. La première, issue de nos travaux de recherche en master, a été réalisée en 2012 par C. Garcia-Debanc. Il s'agit de trois classes de CM2 (CM2 B, CM2 S, CM2 T) et d'une classe de CM1 (CM1 B). Les collectes suivantes ont été effectuées par nos soins durant les deux premières années de thèse. Une collecte en juin 2015 a permis de recueillir des textes d'une classe de CM1 (CM1 T1) et d'une classe de CE2 (CE2 L). Une deuxième collecte en juin 2016, a permis également le recueil des textes d'une classe de CM1 (CM1 T2) et de CE2 (CE2 P). Enfin, un dernier recueil en juin 2017 a permis de réunir des textes d'une classe de CM2 (CM2 M) et de trois classes de CE2 (CE2 C, CE2 T, CE2 U). Cette dernière collecte visait à équilibrer le nombre de textes par niveau scolaire. Nous précisons que certaines classes sont des classes multiniveaux.

Au final, nous avons recueilli **210** textes d'élèves. Le tableau ci-après montre la composition de notre corpus.

Classe	Nombre de textes recueillis par classe	Nombre de textes recueillis par niveau
CE2 C	6	
CE2 L	8	
CE2 P	22	
CE2 T	25	
CE2 U	9	70
CM1 B	15	
CM1 T1	29	
CM1 T2	26	70
CM2 B	15	
CM2 S	25	
CM2 T	21	
CM2 M	9	70
TOTAL	**210**	**210**

Tableau 1 : Nombre de textes recueillis par classes et niveau scolaire

Parmi les textes collectés, nous avons écarté de l'analyse les textes ne répondant pas à la consigne, c'est-à-dire les textes dans lesquels les phrases imposées ne respectent pas l'ordre indiqué ou sont collées les unes à la suite des autres, ainsi que les textes contenant un trop grand nombre d'éléments illisibles, rendant l'exploitation impossible. Ainsi, nous avons pu exploiter 185 textes.

Le corpus est constitué des textes d'élèves scannés et de leurs transcriptions (cf. section 6.3.) ainsi que de métadonnées indiquant l'identité du scripteur, son sexe, sa date de naissance, son niveau scolaire et la date de réalisation de l'exercice.

Mise à part la première collecte, tous les recueils de données ont été précédés de plusieurs documents :

- une note explicative à l'attention des parents expliquant les enjeux de cette recherche
- une autorisation d'exploitation à signer par les parents ou représentants légaux de l'élève. Ce dernier document vise à autoriser l'exploitation des données dans un cadre scientifique plus large que la seule réalisation d'une thèse de doctorat, comme la mise à disposition du corpus pour tout membre de la communauté scientifique.

Les enseignants qui ont accepté de participer à notre projet de recherche sont issus de notre propre réseau de connaissances ou des réseaux professionnels de collègues didacticiens. La composante didactique n'étant pas notre objet d'étude, nous n'avons pas tenu compte des caractéristiques de l'environnement didactique des classes, ni recueilli d'informations concernant les enseignants.

6.3. Méthodologie de travail pour la transcription des textes du corpus

Un des objectifs de notre travail de thèse est de constituer un corpus de textes scolaires exploitable par la communauté scientifique quel que soit l'angle d'étude considéré. Pour cela, il est nécessaire d'envisager une transcription des textes d'élèves qui soit le plus près possible du texte manuscrit. Nous avons donc choisi de transcrire chaque texte de manière la plus exhaustive possible, en mêlant les principes d'une transcription diplomatique (être le plus fidèle au texte possible) et ceux d'une transcription linéarisée (codage typographique de tous les signifiants graphiques présents dans le manuscrit) à l'image du corpus *Ecriscol* (Doquet & *al.*, 2017 ; Testenoire, 2017). Dans un second temps, nous avons réalisé une transcription normée afin que le corpus puisse être analysé *via* n'importe quel outil d'aide à l'annotation.

Cependant, nous l'avons déjà vu, les spécificités des textes d'élèves soulèvent de nombreuses difficultés (cf. chapitre 5) et nécessitent souvent de faire des choix méthodologiques dans les transcriptions. Nous nous proposons ici d'en rendre compte.

6.3.1. Transcription diplomatique

La première transcription réalisée est une transcription diplomatique afin de garder une iconicité maximale des textes du corpus. En l'absence de convention partagée de transcription, nous nous sommes appuyées sur le Projet de Guide de Transcription de Textes d'élèves issu des journées d'études de juin 2015 à Toulouse[53]. Lors de cette rencontre, une

[53] Journées d'études organisées par C. Garcia-Debanc : *Méthodes d'analyse et de traitement des textes d'élèves : enjeux scientifiques et ressources pour la formation*, Toulouse, juin 2015. Ces journées faisaient suite à celle organisée par C. Doquet : *Analyser informatiquement des grands corpus d'écrits scolaires : problèmes de transcription, d'annotation et de traitement*, Paris, mars 2015.

réunion préparatoire au dépôt d'un projet ANR impliquant des collègues des équipes de CLLE-ERSS, Grenoble, Paris3 et Paris 8 a été organisée. Elle avait notamment pour objectif de tenter de définir des conventions de transcriptions de textes d'élèves applicables par tous. De cette réunion est né un projet de guide de transcription ayant le mérite de définir les codages des éléments les plus généraux (cf. sections 6.3.1.1 et 6.3.1.2.), même si des difficultés et interrogations demeurent. Le guide s'inspire notamment des conventions de codage utilisées par l'Institut des Textes et Manuscrits (ITEM), relevant de la génétique textuelle. Beaucoup des conventions mises en place lors de ces journées ont été utilisées dans le projet *Ecriscol* (cf. chapitre 5, section 5.1.2.).

Dans cette première transcription, tous les textes sont reproduits en gardant l'orthographe et la ponctuation initiales de l'élève (sans normalisation) puisque le but est de reproduire le plus fidèlement possible ce qui est visible sur le texte originel. Les phrases imposées par la consigne sont reproduites en italique dans la transcription.

6.3.1.1. Codage des opérations d'écriture : ajout, suppression, remplacement, déplacement et ratures

Les textes d'élèves, de même que les brouillons d'écrivain, sont caractérisés par la présence de ratures, d'ajouts, de suppressions et de remplacements de caractères, de mots ou de phrases. Ces éléments sont d'ailleurs souvent combinés ou accumulés. Un texte peut donc présenter des suppressions ou des ajouts à la fois de la part de l'élève et de celle de l'enseignant (lorsqu'il s'agit par exemple d'un corpus comprenant les annotations de l'enseignant). Ces opérations peuvent être effectuées au moment de l'écriture ou bien ultérieurement, comme le signalerait une couleur d'encre différente par exemple. Le protocole de transcription doit donc permettre un codage clair et accessible de chaque processus (Doquet & *al.*, 2017a et b).

Dans la mesure où notre consigne ne prévoit pas de réécriture du texte produit par les élèves, nous ne sommes pas confrontés aux problèmes soulevés par les différentes « voix » du texte (élève/enseignant) et les différents moments de l'écriture à plusieurs jours d'intervalles, se traduisant par plusieurs versions d'un même texte. Pour autant, la question des différents

temps d'écriture se pose tout de même puisqu'un simple ajout ou une rature peuvent être considérés comme un retour en arrière dans le processus rédactionnel, relevant de la révision (ou *reviewing*, Hayes & Flower, 1981 ; cf. chapitre 3, section 3.1.). Les ratures, ajouts, suppressions, remplacement et déplacement d'éléments du texte sont des marques qui permettent au lecteur de reconstituer les opérations scripturales (Grésillon 1994, cité par Doquet & *al.,* 2017). Pour coder ces opérations lors de la transcription des textes, nous utilisons :

- les chevrons pour marquer l'ajout d'un segment de texte : **<segment ajouté>**
- les crochets pour marquer la suppression : **[segment supprimé]**
- une combinaison du signe dièse et de x pour marquer un segment illisible, quelle que soit sa taille : **#xxx#** lorsque le segment est isolé, **#x#** lorsque le segment est compris dans un mot lisible
- une combinaison des codes décrits ci-dessus lorsqu'il s'agit de la suppression d'un segment non lisible (trace de gomme, d'effaceur, de blanc ou rature) : **[#xxx#]**
- Le codage des opérations de remplacement et de déplacement est la combinaison des opérations d'ajout et de suppression, avec un signe distinctif (@1) identifiant le segment déplacé

Nous illustrons ces codages par la reproduction ci-dessous d'un extrait de texte d'élève (6) suivi de sa transcription.

abitait dans un manoir entée ils avai#x#ntune voisine.

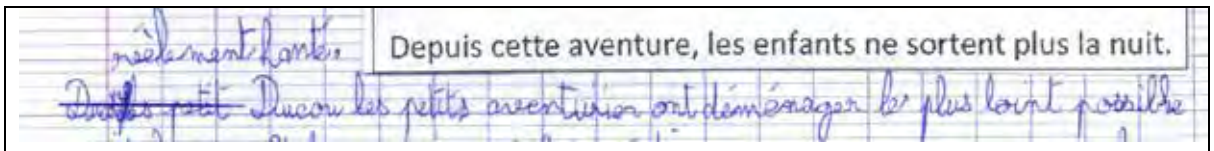
Elle habitait dans cette maison depuis longtemps. Un jour [Tom] en hiver <pendant la nuit> Tom se réveilla en surso acause d'un grand bruit #xxx# il entandit Chou ! #xxx# de la fumé sortit de la sale à manger.Le lendemain

Il se retourna en entendant ce grand bruit. [Cette] Ce manoir était bien et

Figure 6 : Texte de Nina, CE2 L

A la troisième ligne du texte, un segment non lisible est présent entre les mots *bruit* et *il*. Il est donc codé par trois caractères x entre dièses : bruit #xxx# il. A la première ligne reproduite du texte de l'élève, il y a un segment non lisible au milieu du verbe *avoir,* codé par un seul caractère x entre dièses : avai#x#ntune. L'étude très détaillée de la copie (zoom plusieurs fois) montre qu'il s'agit probablement de la lettre *e* sur laquelle l'élève a inscrit la lettre *t* puis l'a biffée, signe d'une hésitation entre forme verbale au pluriel et forme verbale au singulier. Cet élément pose une question difficile et caractéristique de la transcription : jusqu'à quel degré de détail doit-on aller dans la transcription ? À partir de quel moment y a-t-il basculement de la transcription vers une analyse interprétative ? Nous avons fait le choix de ne transcrire que les éléments visibles à l'œil nu (sans zoom sur la copie) et ne nécessitant pas d'étude approfondie. Le codage utilisé suffit à marquer la présence d'un élément remarquable, que chacun sera libre d'étudier plus en détail en revenant au scan du texte. Par exemple, la trace graphique sur la troisième ligne, entre *Chou !* et *de* est codée comme un segment illisible car il peut potentiellement intéresser les généticiens du texte : il pourrait s'agir d'une hésitation de la part de l'élève concernant une trace de guillemets, comme on apprend à l'école qu'une onomatopée s'écrit entre guillemets.

Le souci de reproduire le plus fidèlement possible les éléments visibles sur le texte de l'élève peut nous amener à combiner les codages, comme par exemple dans (7) :



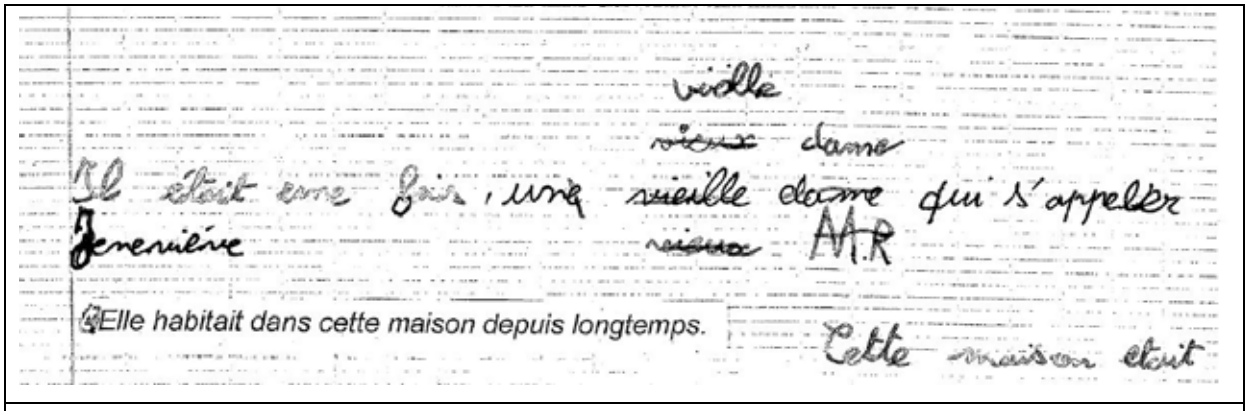
réèlement hanté. *Depuis cette aventure, les enfants ne sortent plus la nuit.*
[<Duco>#x#Les petit] Ducou les petits aventurier ont déménager le plus loint possible

Figure 7 : Texte de Nina, CE2 L

Au début de la deuxième ligne, l'élève a ajouté un segment dans la marge, codé ainsi entre chevrons. Ce segment se termine par un élément non lisible, codé par #x#. Mais cet ajout est constitutif d'un segment plus long, lui-même finalement supprimé, donc codé entre crochets.

La transcription relativement simple de ces combinaisons d'opérations scripturales trouve ses limites dans les textes fortement remaniés, comme dans (8) ci-dessous. Comme il s'agit de plusieurs opérations de remplacement d'éléments par d'autres, il faudrait coder ces

opérations par des suppressions entre crochets puis des ajouts entre chevrons. Mais dans cet extrait, il est impossible de savoir dans quel ordre se sont succédées les biffures (d'abord celle au-dessus de la ligne ou celles en-dessous ?). Dans ces cas-là, nous avons choisi une transcription qui respecte l'utilisation de l'espace de la page par l'élève et qui se contente de marquer les suppressions, sans risquer une interprétation concernant les moments auxquels interviennent les différents ajouts.



vieille

[vieux] dame

Il était une fois, un[e] [vieille] [dame] qui s'appeler

Jeneviève [vieux] [MR]

Elle habitait dans cette maison depuis longtemps. Cette maison était

Figure 8 : Texte de Samuel, CM2 B

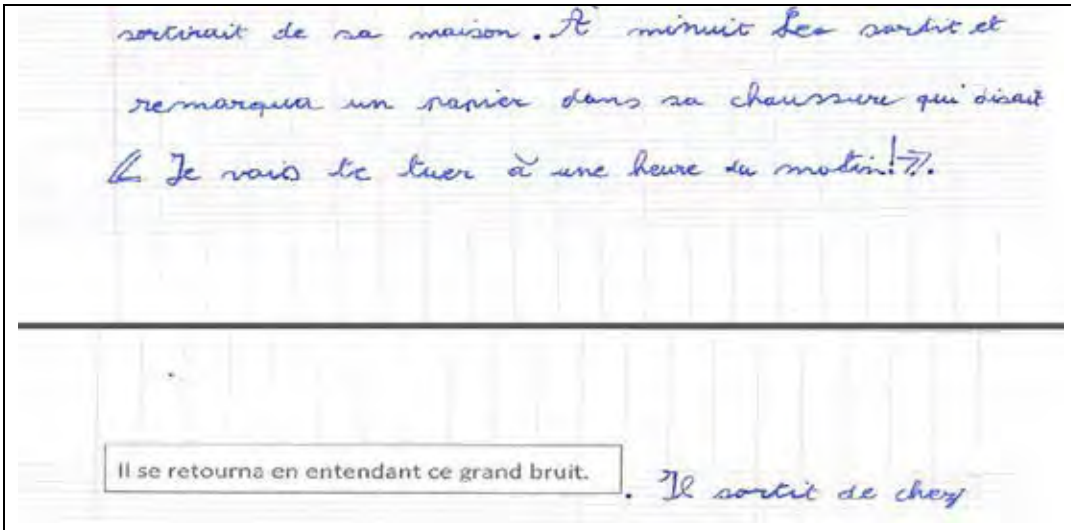
Ces très nombreuses ratures témoignent d'une hésitation sur le genre du personnage. Il semble que l'élève a d'abord introduit un personnage féminin mais qu'il a ensuite rectifié en masculin, sans doute en essayant d'intégrer la deuxième phrase.

6.3.1.2. Codage de la mise en page

Dans l'écriture, la contrainte du support est importante. Aussi, la mise en page du texte a également été codée afin d'être transcrite. Nous avons repris les conventions définies dans le projet de guide de transcription de textes d'élèves :

- le symbole **§** est inscrit en fin de ligne lorsque l'élève effectue un retour à la ligne volontaire
- le symbole **§** est inscrit sur la ligne laissée vierge lorsque l'élève saute une ligne
- les caractères **\n** sont inscrits en fin de ligne lorsque l'élève effectue un retour à la ligne imposé par une frontière de page
- les caractères **\t** sont inscrits en début de ligne lorsque l'élève effectue un alinéa en début de paragraphe

Nous illustrons ce codage par la reproduction ci-dessous d'un extrait de texte d'élève suivi de sa transcription.



sortirait de sa maison. À minuit Léo sortit et

§

remarqua un papier dans sa chaussure qui disait

§

« Je vais te tuer à une heure du matin ! ». §

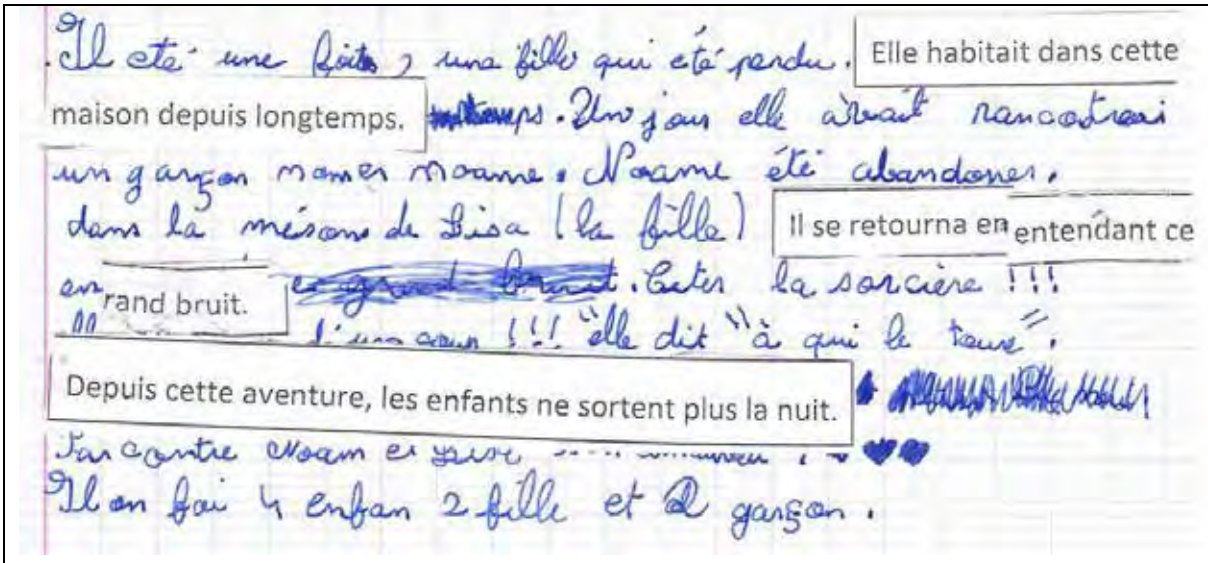
§ \n

Il se retourna en entendant ce grand bruit. Il sortit de chez

Figure 9 : Texte de Martin, CM1 T2

Bien que la reproduction du texte la plus fidèle possible soit l'objectif de notre transcription, il reste des éléments difficilement intégrables pour diverses raisons. La

difficulté vient parfois du fait que leur reproduction exacte n'est pas envisageable via un outil informatique, comme c'est le cas par exemple des dessins (deux petits cœurs sur l'avant-dernière ligne) du texte ci-dessous (10). Dans d'autres cas, la multiplication des codages rendrait la tâche de transcription complexe et coûteuse et la lecture fastidieuse. C'est également le cas lorsqu'il s'agit « d'accident » de mise en page, comme la bandelette de papier qui se déchire (5e ligne) ou qui est collée en partie sur le texte (6e ligne). Dans ces cas-là, nous avons choisi de transcrire uniquement les segments visibles et d'indiquer la présence de ces éléments par un commentaire libre dans l'en-tête de notre fichier de transcription, juste avant la transcription du texte correspondant. Cela permet d'avertir le lecteur de l'existence d'éléments non transcrits.



Il été une foi#x#, une fille qui été perdu. *Elle habitait dans cette maison depuis longtemps.* [#x#tem]ps. Un jour elle àvait rancotrai un garçon nomer noame. Noame été abandoner. §
dans la méson#x# de Lisa (la fille) *Il se retourna en entendant ce* en *rand bruit*. [e grand bruit]. Ceter la sorciere !!! §
[#xxx#] d'un coup !!! « elle dit « à qui le toure ». §
Depuis cette aventure, les enfants ne sortent plus la nuit. [#xxx#] [#xxx#]
Par contre Noam et Lisa [#xxx#]
Il on fai 4 enfan 2 fille et 2 garçon.

Figure 10 : Texte de Loulie, CE2 P

6.3.1.3. Récapitulatif des conventions de codage utilisées pour la transcription diplomatique

Nous reproduisons dans le tableau ci-dessous les conventions de codage utilisées pour la transcription diplomatique de notre corpus.

Éléments à transcrire	Codage
Ajout de texte	<segment ajouté>
Suppression de texte	[segment supprimé]
Déplacement de texte	[segment déplacé]@1 puis < >@1 à l'endroit ciblé
Segment illisible isolé	#xxx#
Segment illisible à l'intérieur d'un autre segment	#x#
Suppression d'un segment illisible	[#xxx#]
Retour à la ligne volontaire	§ *en fin de ligne*
Retour à la ligne imposé par l'espace de la page	\n *en fin de ligne*
Saut de ligne	§ *sur la ligne sautée*
Alinéa	\t *en début de ligne*

Tableau 2 : Synthèse des conventions de codage utilisées pour la transcription diplomatique du corpus

Tout ce qui n'est pas présent dans le tableau (dessins, accidents de mise en page, etc.) n'a pas été codé dans la transcription mais fait l'objet d'un signalement sous la forme d'un commentaire libre dans l'en-tête du fichier de transcription de chaque texte concerné.

6.3.2. Transcription normée

Chaque texte a fait l'objet d'une deuxième transcription ayant pour objectif le rétablissement de l'orthographe et de la ponctuation. En effet, les textes d'apprentis scripteur se caractérisent par leur écart à la norme orthographique et ponctuationnelle (cf. chapitre 5).

Cet écart pose problème lorsqu'il s'agit de traiter les textes automatiquement, c'est-à-dire par le biais d'outils d'aide à l'annotation qui, pour la plupart, ne sont pas adaptés à ce type de textes. Toutefois, la normalisation fait perdre de vue les graphies et choix originels du scripteur (Boré & Elalouf, 2017). C'est d'ailleurs pour pallier ce problème que nous avons réalisé une première transcription qui se veut diplomatique.

Bien que certains travaux aient réussi à utiliser une graphie acceptable pour le logiciel d'analyse sans faire disparaître la graphie originelle de l'élève (cf. corpus *Ecriscol*), cela n'a pas été notre choix. Notre objet d'étude ne portant pas sur le traitement systématique des erreurs (d'orthographe, de conjugaison…) ou la ponctuation mais sur la cohérence et la cohésion textuelle, nous avions besoin d'une version « nettoyée » afin que l'orthographe et la ponctuation ne soient pas des obstacles supplémentaires à la lecture des textes, déjà difficilement compréhensibles par ailleurs pour certains. L'effacement des graphies qui perturbent les habitudes de lecture permet de se centrer sur la textualité (Boré & Elalouf, 2017). En outre, la version normée permet d'être lue par le logiciel d'aide à l'annotation de la référence que nous avons sélectionné (nous reviendrons sur ce choix plus loin).

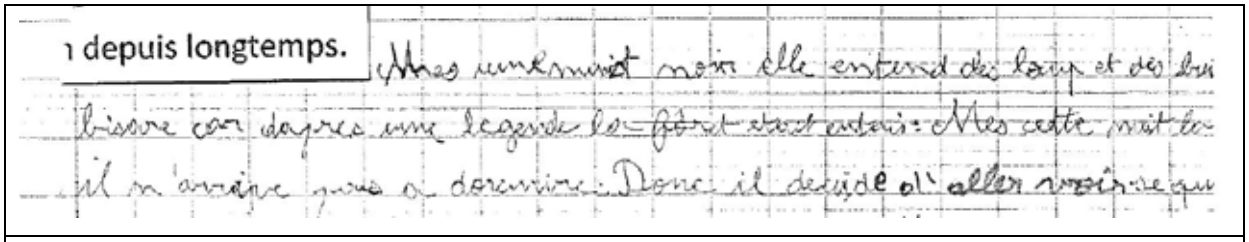
Que ce soit le rétablissement de la norme orthographique ou celui d'une ponctuation normée, chaque type de transcription est déjà une forme d'interprétation du texte et donc d'analyse, comme nous nous apprêtons à le montrer. Les choix effectués pour la normalisation des textes de notre corpus relèvent eux-aussi d'une forme d'analyse initiale et ne sont donc pas exploitables tels quels par des chercheurs dont l'objet d'étude diffère du nôtre. Pour cette raison, le premier type de transcription que nous avons effectuée, diplomatique, est indispensable car elle permet de revenir au texte originel.

6.3.2.1. Normalisation orthographique

Toute transcription étant déjà le résultat de choix de transcripteur, la version orthographiée selon la norme des textes a fait elle aussi l'objet de différents choix, qui ne se résument pas uniquement à des choix syntaxiques.

Choix des pronoms

Le tout premier concerne les pronoms. En effet, les apprentis scripteurs les plus jeunes ont tendance à employer des pronoms masculins à la place de pronoms féminins (Cappeau & Roubaud, 2005), et cela même s'ils ont préalablement mentionné une occurrence de genre féminin, comme dans (11) ou *elle* devient *il* à la troisième ligne :



depuis longtemps. Mes une nui[s]<t> noir elle entend des loup et des bui
bisare car dapres une legende la forêt était entais : Mes cette nuit la
il n'arrive pas a dor#x#mire : Donc il decide d'aller voir se qu

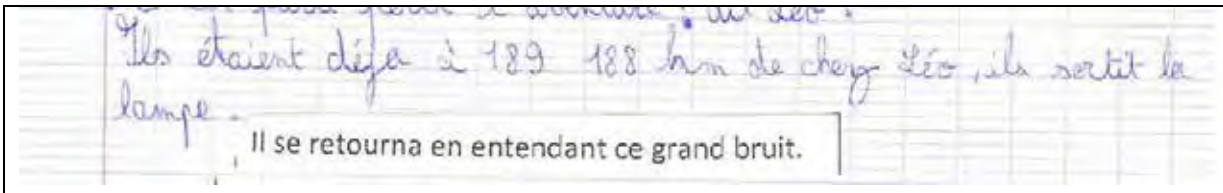
depuis longtemps. Mais une nuit noire, elle entend des loups et des bruits bizarres car d'après une légende, la forêt était hantée. Mais cette nuit-là, il n'arrive pas à dormir. Donc il décide d'aller voir

Figure 11 : Texte de Calixte, CM1 T1

Si dans la version normée, le pronom masculin est remplacé par un pronom féminin, c'est la cohérence du texte qui est corrigée. Or c'est précisément l'objet de notre étude. Nous avons donc décidé de garder tels que les élèves les ont rédigés tous les éléments relevant de la cohérence et la cohésion textuelle des occurrences que nous étudions.

Dans l'extrait (12), la phrase débute avec un pronom à la troisième personne du pluriel, *ils,* et le verbe *être* est correctement accordé. Mais dans la deuxième partie de la phrase, le pronom pluriel *ils* est suivi d'un verbe au singulier : *sortit*. Sur quel mot s'appuyer pour rétablir une version normée, le pronom ou le verbe ? Le choix de l'un ou de l'autre n'est pas anodin car il peut modifier la cohérence locale. En effet, si on garde le pronom au pluriel et qu'on modifie le verbe en conséquence, *ils sortirent la lampe*, le lecteur peut avoir du mal à identifier le référent du pronom personnel singulier de la phrase suivante. Si on modifie le pronom pluriel *ils* en pronom singulier *il*, *il sortit la lampe,* le calcul référentiel serait alors plus facile. Comme il est impossible de connaître le raisonnement de l'élève qui a abouti à ces graphies, dans des cas comme celui-ci, nous avons fait le choix de donner la priorité au

pronom et de rétablir le verbe en conséquence car nous considérons que l'élève n'a pas de connaissance morphologiques suffisantes du passé simple.



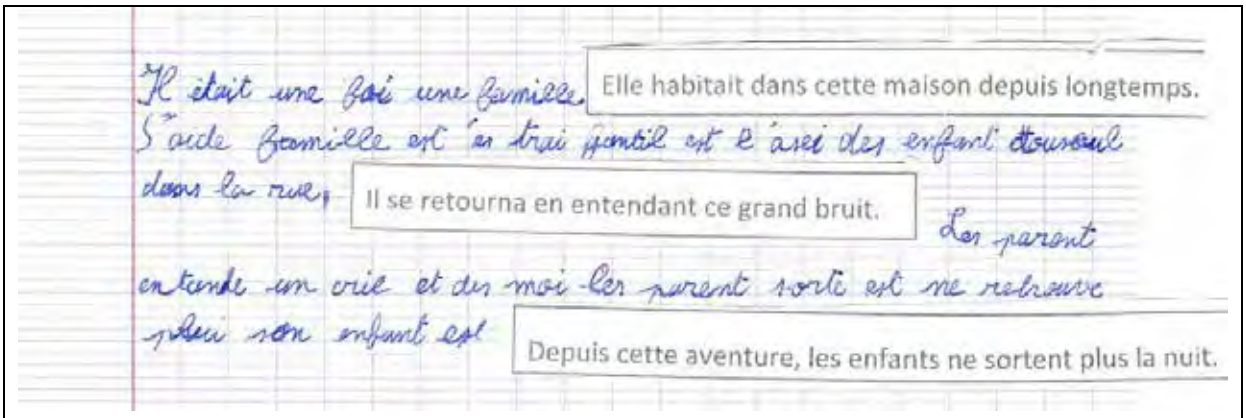
Ils étaient déja à 189 188 km de chez Léo, ils sortit la
lampe.
Il se retourna en entendant ce grand bruit.

Ils étaient déjà à 189 188 km de chez Léo, ils sortirent la lampe. *Il se retourna en entendant ce grand bruit.*

Figure 12 : Texte d'Agathe, CE2 P

Choix de l'orthographe

En ce qui concerne l'orthographe, la maîtrise plus ou moins grande des scripteurs donne lieu à des textes parfois très difficiles à déchiffrer. Pour comprendre ce genre de textes, nous recourons à la lecture à voix haute afin de redonner du sens et de la cohérence au texte de l'élève, à l'instar notamment de Fabre-Cols (2000) et Roubaud (2017). Ainsi, un texte difficilement compréhensible comme (13) a pu être rétabli.



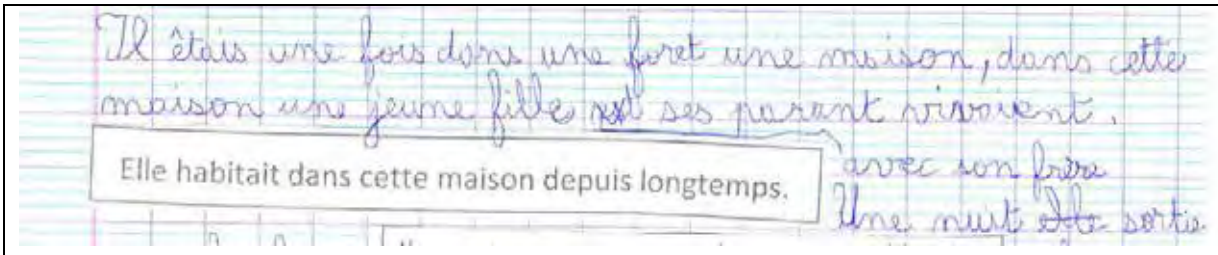
Il était une fois une famille. *Elle habitait dans cette maison depuis longtemps.* Cette famille était très gentille et avait des enfants tous seuls dans la rue. *Il se retourna en entendant ce grand bruit.* Les parents entendent un cri : « aidez-moi ! ». Les parents sortent et ne retrouvent plus leur enfant et *depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 13 : Texte de Florentin, CE2 T

Cependant, certains cas de figure résistent. Par exemple, que faire de *l'asei* à la troisième ligne ? L'élève a-t-il voulu écrire *l'avait, elle avait, laissait* ou autre chose encore ? Dans ces cas-là, le choix se porte sur le sens le plus logique.

Choix de l'ordre des mots

Rétablir la norme orthographique des textes scolaires peut amener à s'interroger sur le rétablissement de l'ordre des mots. Dans (14) par exemple, on comprend à la première lecture qu'une jeune fille vit dans une maison avec ses parents et son frère, malgré le fait que le verbe est en fin de phrase […] *dans cette maison, une jeune fille avec son frère et ses parents vivaient.* Pourtant, remanier l'ordre des mots dans la première phrase en déplaçant le verbe rend plus confortable la lecture et la compréhension :

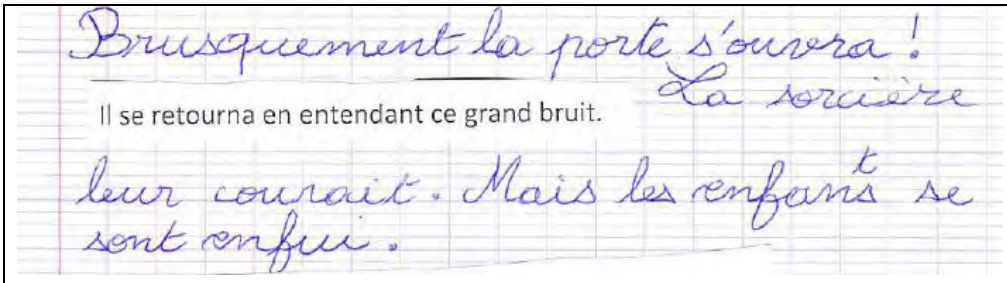


Il était une fois dans une forêt, une maison, dans cette maison, une jeune fille vivait avec son frère et ses parents. *Elle habitait dans cette maison depuis longtemps.*

Figure 14 : Texte de Karelle, CE2 T

Toutefois, le remaniement de l'ordre des mots dans cet exemple n'est pas indispensable sur le plan sémantique : quelle que soit la place du verbe, on comprend la phrase. Cet exemple illustre le fait qu'il est difficile de déterminer la frontière entre un rétablissement qui rend le texte seulement plus confortable à lire et un rétablissement vraiment nécessaire.

En outre, le rétablissement des constructions syntaxiques nécessite parfois d'ajouter des mots qui n'apparaissent pas dans le texte initial, comme par exemple la préposition *après* (*courait après*) dans le (15) ci-dessous :

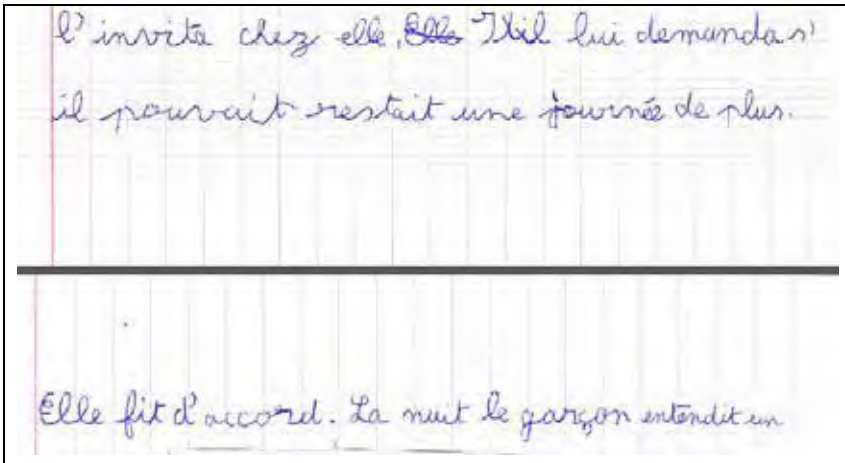

Brusquement la porte s'ouvra ! La sorcière
Il se retourna en entendant ce grand bruit.
leur courait. Mais les enfants se
sont enfui.

Brusquement, la porte s'ouvrit ! *Il se retourna en entendant ce grand bruit.* La sorcière leur courait après. Mais les enfants se sont enfuis.

Figure 15 : Texte d'Aurore, CE2 P

Nécessité d'interprétation

Enfin, la normalisation orthographique nécessite parfois de faire des choix d'interprétation. Considérons l'extrait (16) suivant :


l'invita chez elle. Il lui demanda s'
il pouvait restait une journée de plus.
Elle fit d'accord. La nuit le garçon entendit un

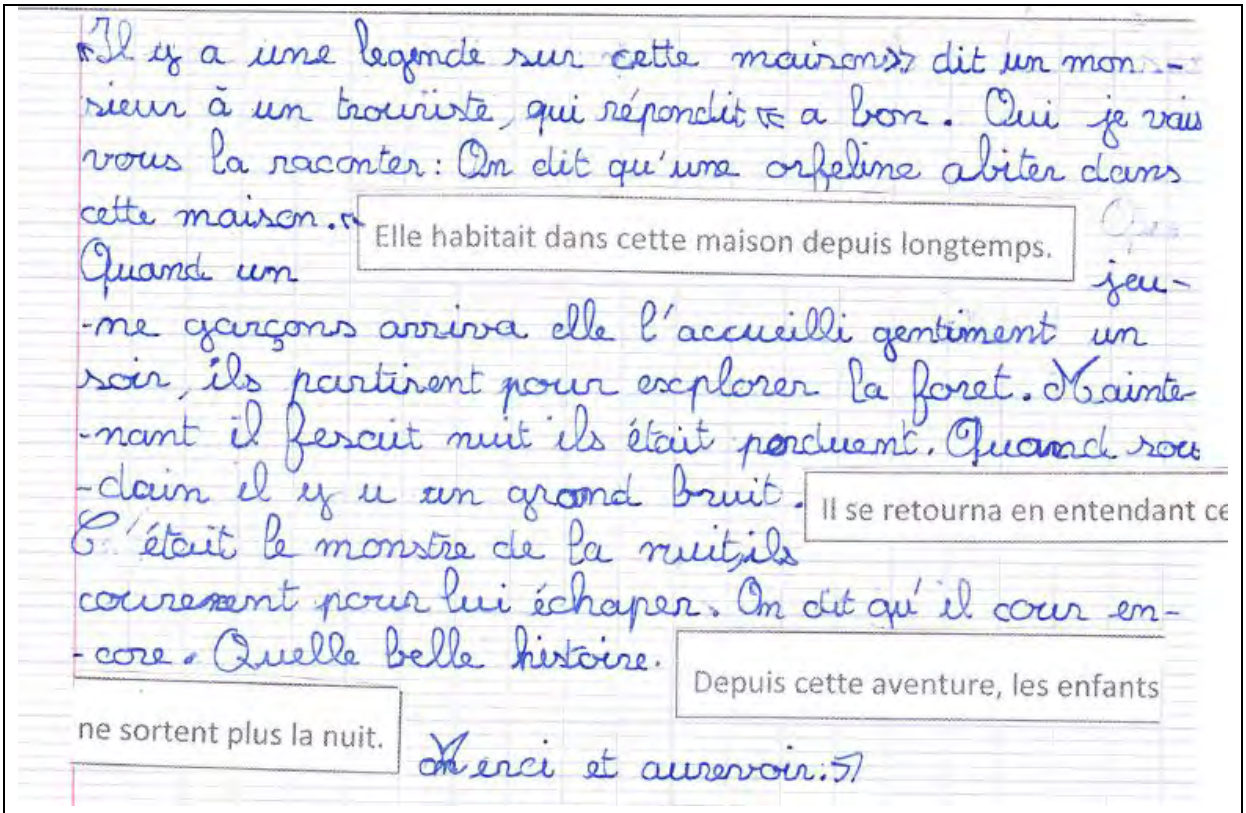
l'invita chez elle. Il lui demanda s'il pouvait rester une journée de plus. Elle fut d'accord. La nuit le garçon entendit un

Figure 16 : Texte de Romain, CE2 P

Dans cet extrait, on peut considérer que le personnage féminin est d'accord et le signifie en disant « d'accord », mais que l'élève a oublié la ponctuation spécifique au dialogue. Dans ce cas, la phrase sera rétablie de la manière suivante : *Elle fit : « D'accord ».* Mais l'on peut

aussi supposer que l'absence de guillemets signifie que l'élève n'a pas voulu intégrer un élément de dialogue à cet endroit mais qu'il a plutôt fait une erreur de graphie et a voulu écrire *Elle fut d'accord*.

De la même manière, dans (17) il est difficile de distinguer le discours du monsieur et celui du touriste.


Il y a une legende sur cette maison dit un mon-
sieur à un touriste, qui répondit a bon. Oui je vais
vous la raconter : On dit qu'une orfeline abiter dans
cette maison.
Elle habitait dans cette maison depuis longtemps.
Quand un jeu-
-ne garçons arriva elle l'accueilli gentiment un
soir, ils partirent pour explorer la foret. Mainte-
-nant il fesait nuit ils était perduent. Quand sou
-dain il y u un grand bruit.
Il se retourna en entendant ce
C'était le monstre de la nuit, ils
coureserent pour lui échaper. On dit qu'il cour en-
-core. Quelle belle histoire.
Depuis cette aventure, les enfants
ne sortent plus la nuit.
Merci et aurevoir.

« - Il y a une légende sur cette maison », dit un monsieur à un touriste, qui répondit : « Ah bon ».

- Oui, je vais vous la raconter : on dit qu'une orpheline habitait dans cette maison. « *Elle habitait dans cette maison depuis longtemps.* Quand un jeune garçon arriva, elle l'accueillit gentiment. Un soir, ils partirent pour explorer la forêt. Maintenant, il faisait nuit. Ils étaient perdus. Quand soudain, il y eut un grand bruit. *Il se retourna en entendant ce grand bruit.* C'était le monstre de la nuit. Ils coururent pour lui échapper. On dit qu'il court encore.

- Quelle belle histoire !

- *Depuis cette aventure, les enfants ne sortent plus la nuit.*

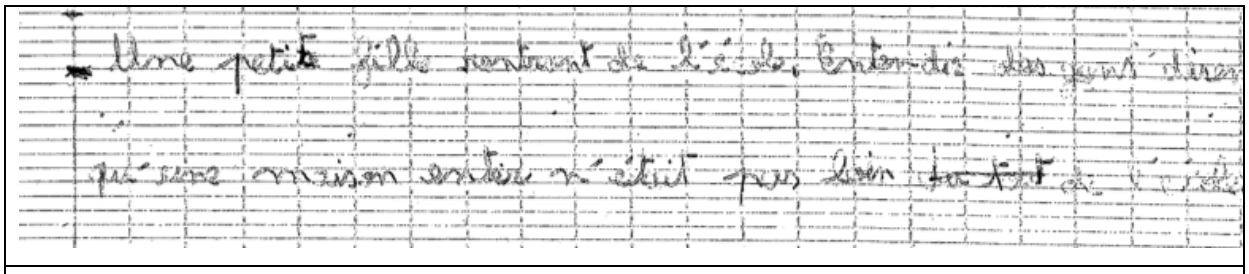
- Merci et au revoir».

Figure 17 : Texte de Juliette, CE2 T

La dernière phrase : *Merci et au revoir* peut être attribuée tout autant à l'un des personnages qu'à l'autre mais cela ne sera pas rendu de la même manière dans la version normalisée. Encore une fois, l'interprétation la plus logique est privilégiée : il semble plus naturel que ce soit le touriste qui remercie le conteur de lui avoir communiqué cette histoire, plutôt que le conteur qui remercie le touriste de l'avoir écouté.

6.3.2.2. Normalisation ponctuationnelle

Les textes d'apprentis scripteurs posent le problème de la construction de la phrase. On rencontre souvent chez les scripteurs les plus jeunes deux types de phrases (Paolacci & Rossi-Gensane, 2012). Le premier est la phrase graphique, qui se traduit par un signe de ponctuation placé entre deux mots sans tenir compte du sens de la phrase. Le deuxième est la phrase syntaxique qui elle, tient compte du sens du propos pour placer la ponctuation. On trouve une illustration de la phrase graphique dans (18) ci-dessous, où l'élève a mis un point après *école* sur la première ligne.



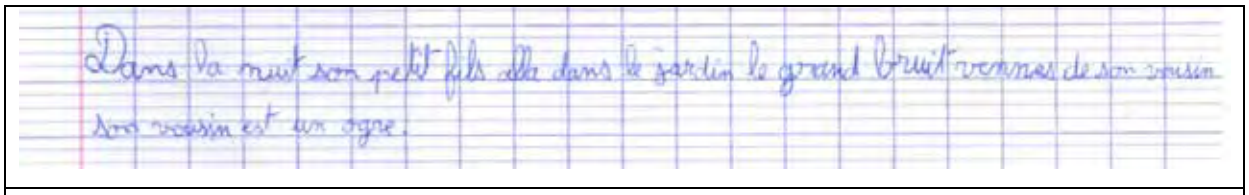
Une petite fille rentrant de l'école. Entendis des gens direns
§
qu'une maison enter n'était pas loin [du tout] de l'école.

Figure 18 : Texte de Jules, CM2 B

Les élèves ont souvent tendance à recourir à la phrase graphique dans leur texte, donnant lieu parfois à une « sur-ponctuation ». Ce procédé est utilisé par certains auteurs pour produire un effet de rythme mais les élèves scripteurs ne sont pas encore capables de maîtriser ces effets (Cappeau & Roubaud, 2005). Il s'agit alors de la part des élèves d'une confusion au niveau de la fonction démarcative de la phrase (assurer le découpage du texte en unités) sur

les plans graphiques et syntaxiques. L'extrait ci-dessus illustre d'ailleurs bien ce phénomène : l'élève met un point après un segment décrivant une action au participe présent, sans doute afin de le séparer du segment d'après qui, lui, décrit une action au passé simple, confondant ainsi les plans graphiques et syntaxiques.

L'effet inverse, la « sous-ponctuation », est également présent dans les textes des apprentis scripteurs, particulièrement chez les plus jeunes, comme le montre (19).

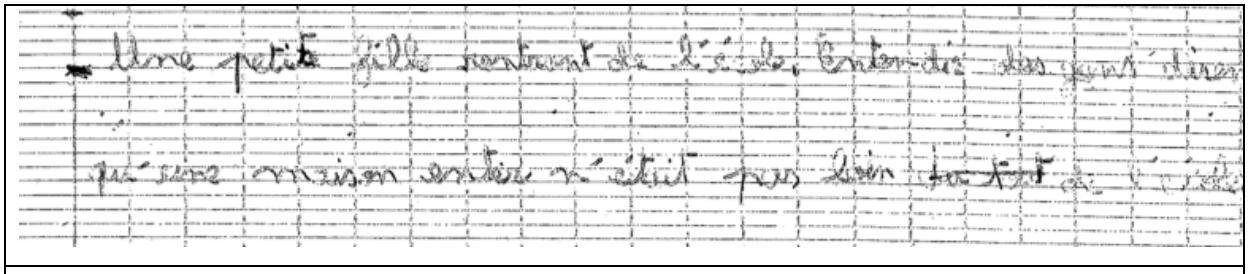


Dans la nuit son petit fils alla dans le jardin le grand bruit venait de son voisin
son voisin est un ogre.

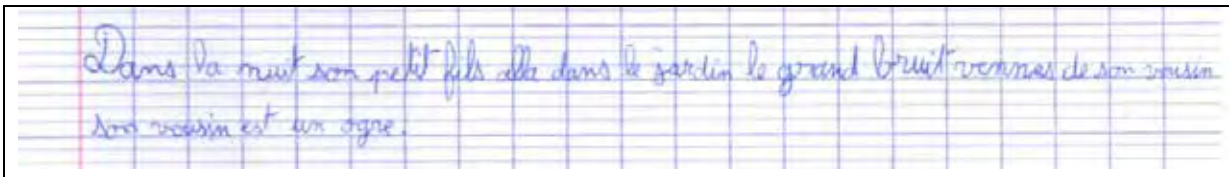
Figure 19 : Texte de Laly, CE2 L

Quel que soit le mode de ponctuation employé par l'élève (sur-ponctuation, sous-ponctuation ou ponctuation normée, phrase graphique ou phrase syntaxique), elle traduit la délimitation de blocs syntaxiques et énonciatifs qui font sens pour lui (Boré & Elalouf, 2017). C'est un moyen pour le chercheur qui s'intéresse à cette question de saisir les représentations de l'élève et ses conceptions de la phrase (Paolacci & Rossi-Gensane, 2012).

Pour la transcription normée des textes présentant une sur-ponctuation ou une sous-ponctuation, nous avons pris la phrase syntaxique comme repère afin de rétablir la norme, comme illustré dans les deux extraits déjà analysés repris ci-dessous.

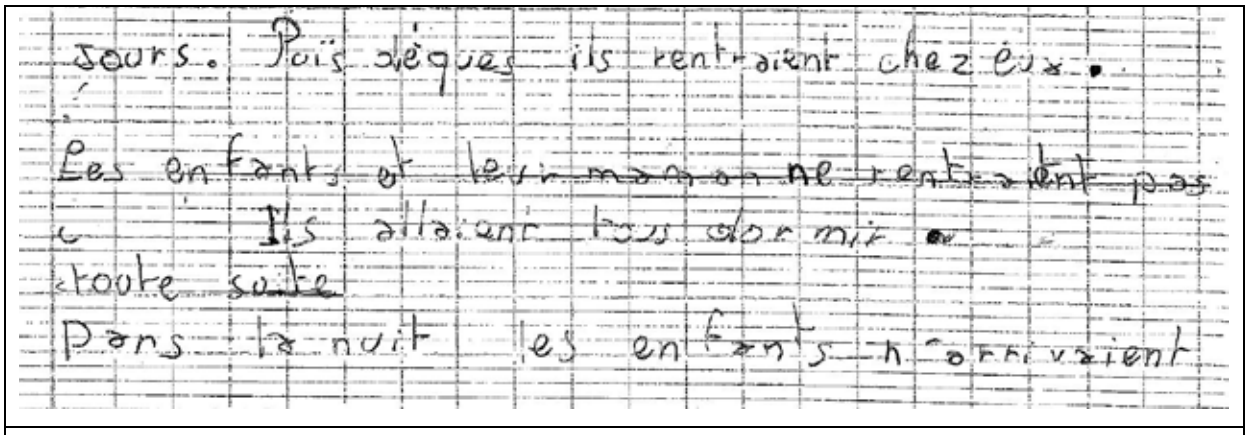


Une petite fille rentrant de l'école entendit des gens dire qu'une maison hantée n'était pas loin de l'école.



Dans la nuit, son petit-fils alla dans le jardin. Le grand bruit venait de son voisin. Son voisin est un ogre.

En outre, les apprentis scripteurs confondent souvent le point et la virgule comme jalons de phrases. Le point est le premier signe de ponctuation enseigné à l'école primaire, puisqu'il apparaît dans les programmes dès le cycle 2 (Cours Préparatoire) et que l'on apprend aux élèves de l'école primaire « qu'une phrase commence par une majuscule et se termine par un point ». Il est, par là-même, le signe de ponctuation majoritairement utilisé par les élèves (Paolacci & Favart, 2010). L'utilisation de la virgule est enseignée plus tardivement : elle apparaît généralement dans les programmes au Cours Élémentaire 2. Mais selon Favart & Chanquoy (2007 : 54), citant Simons (1966), « ces élèves de 8-9 ans ne différencient pas la virgule et le point comme introduisant un degré de rupture différent. Il en résulte une utilisation ambivalente, voire concurrentielle de ces deux signes » (voir aussi à ce sujet Chanquoy & Fayol, 1991 ; Schneuwly, 1988). En effet, le point est le marqueur des constituants narratifs au plan macrotextuel, alors que la virgule est caractéristique du plan microtextuel. Les élèves ne saisissant pas encore la différence entre ces deux signes produisent des textes pour lesquels la lecture - et parfois la compréhension - deviennent difficiles, le lecteur n'arrivant pas à hiérarchiser l'information, comme le montre (20) ci-dessous. Le rétablissement de la norme ponctuationnelle est alors nécessaire : le point après *chez eux* doit être remplacé par une virgule afin de relier ce segment de phrase à la proposition *ils allaient tous dormir*.



Jours. Puis dèques ils rentraient chez eux.

§

[Les enfants et leur maman ne rentrai[e]<t>nt pas]

[#xxx#] <Ils allaient tous dormir.>

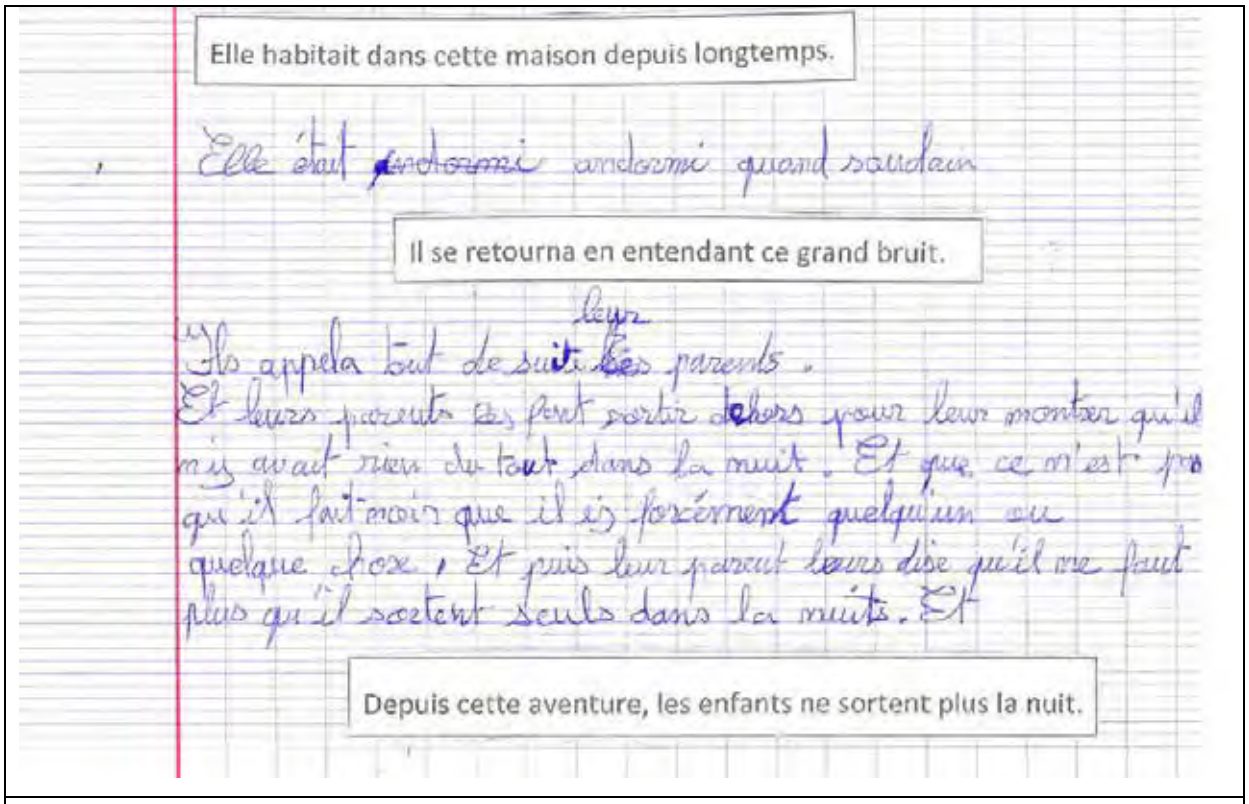
[toute suite]

Dans la nuit les enfants n'arrivaient

jours. Puis dès qu'ils rentraient chez eux, ils allaient tous dormir. Dans la nuit, les enfants n'arrivaient

Figure 20 : Texte de Maëva, CM2 B

Une autre difficulté rencontrée par les apprentis scripteurs est la maîtrise des frontières de phrases. Dans la tâche-problème proposée, trois phrases imposées doivent être insérées dans le texte rédigé par les élèves. Chacune de ces trois phrases est indépendante et n'a pas vocation à être intégrée dans une autre phrase. Pourtant, des élèves n'ont pas tenu compte de la ponctuation et les ont intégrées à leurs propres phrases, comme dans (21) (suivi de sa transcription normée) :



Elle habitait dans cette maison depuis longtemps.

Elle était ~~endormi~~ endormi quand soudain

Il se retourna en entendant ce grand bruit.

Ils appela tout de suite leur parents.
Et leurs parents les font sortir dehors pour leur montrer qu'il n'y avait rien du tout dans la nuit. Et que ce n'est pas qu'il fait noir que il y forcément quelqu'un ou quelque chose, Et puis leur parent leurs dise qu'il ne faut plus qu'il sortent seuls dans la nuits. Et

Depuis cette aventure, les enfants ne sortent plus la nuit.

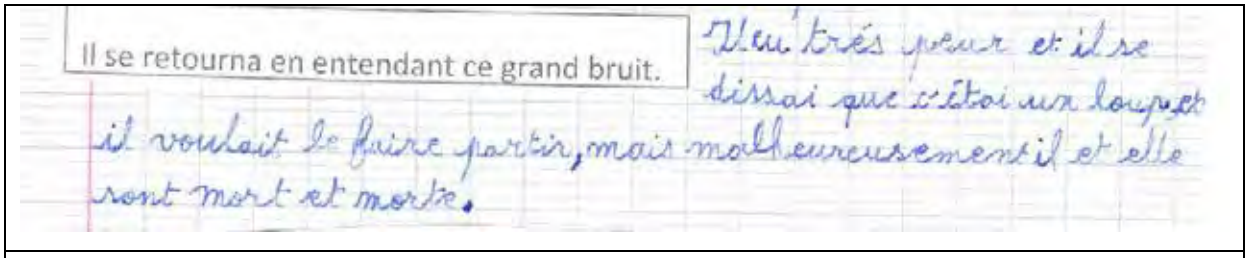
Elle habitait dans cette maison depuis longtemps. Elle était endormie quand soudain, il se retourna en entendant ce grand bruit. Ils appelèrent tout de suite leurs parents. Et leurs parents les font sortir dehors pour leur montrer qu'il n'y avait rien du tout dans la nuit. Et que ce n'est pas parce qu'il fait noir qu'il y a forcément quelqu'un ou quelque chose. Et puis leurs parents leurs disent qu'il ne faut plus qu'ils sortent seuls dans la nuit. Et depuis cette aventure, les enfants ne sortent plus la nuit.

Figure 21 : Texte de Sérine, CM1 T2

Sur cette copie, l'élève a voulu « raccrocher » ensemble les segments *Elle était endormie quand soudain* et la phrase imposée n°2, à travers la présence du syntagme prépositionnel *quand soudain* mais aussi par l'absence de ponctuation entre les deux segments. De la même manière, la phrase imposée n°3 a été raccordée au texte rédigé par l'élève par la présence du connecteur *et*. Lors du rétablissement de la norme ponctuationnelle, nous avons respecté le choix des élèves en intégrant les phrases imposées au texte rédigé par eux, même si cela allait à l'encontre de la consigne qui ne demandait aucune modification des phrases.

Les connecteurs sont fréquents chez les apprentis scripteurs, en particulier le connecteur temporel *et* (très présent à l'oral), qui assure le découpage du texte en unités, souvent à la

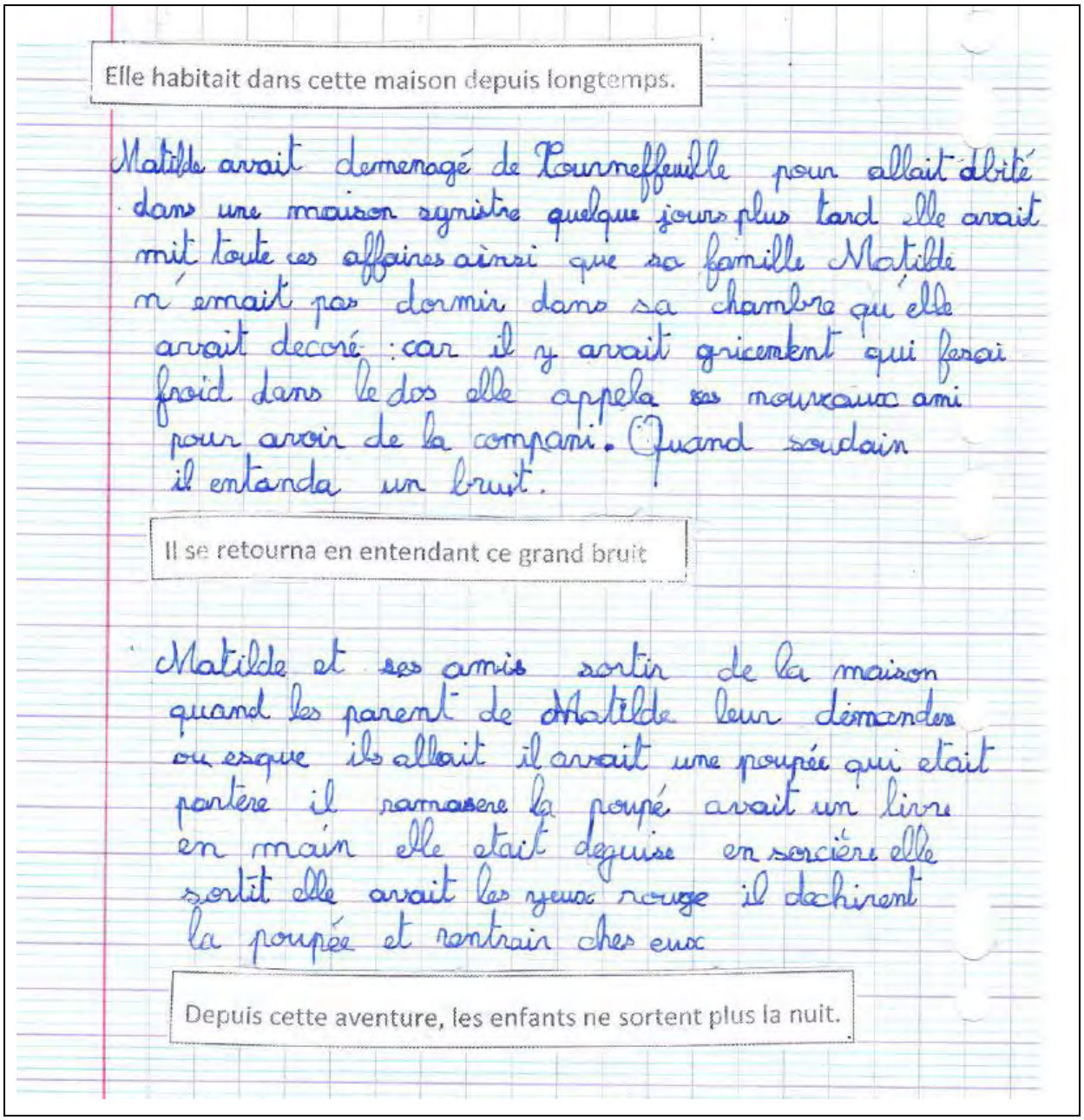
place ou en plus de la ponctuation (Cappeau & Roubaud, 2005). Lorsqu'un texte présente une sur-représentation de ce connecteur, comme dans (22), nous avons fait le choix de remplacer certains *et* par de la ponctuation.



Il se retourna en entendant ce grand bruit. Il eut très peur : il se disait que c'était un loup et il voulait le faire partir, mais malheureusement, il et elle sont mort et morte.

Figure 22 : Texte de Karl, CE2 T

La normalisation de la ponctuation des textes d'élèves soulève aussi des questions d'interprétations et nécessite de faire des choix. Le texte suivant, sous-ponctué, illustre plusieurs des difficultés rencontrées.



Elle habitait dans cette maison depuis longtemps.

Matilde avait demenagé de Tournefeuille pour allait habité dans une maison synistre quelque jours plus tard elle avait mit toute ses affaires ainsi que sa famille Matilde n'emait pas dormir dans sa chambre qu'elle avait decoré : car il y avait gricenkent qui ferai froid dans le dos elle appela ses nouveaux ami pour avoir de la compani. Quand soudain il entanda un bruit.

Il se retourna en entendant ce grand bruit

Matilde et ses amis sortir de la maison quand les parent de Matilde leur demander ou esque ils allait il avait une poupée qui etait partère il ramassere la poupé avait un livre en main elle etait deguise en sorcière elle sortit elle avait les yeux rouge il dechirent la poupée et rentrain chez eux

Depuis cette aventure, les enfants ne sortent plus la nuit.

Figure 23 : Texte de Rayan, CM1 T2

Tout d'abord, dans le segment *Mathilde n'aimait pas dormir dans sa chambre qu'elle avait décoré : car il y avait…* (lignes 5 à 7 de la copie), l'élève a mis deux points suivis de la conjonction de coordination *car*. Cependant, un seul de ces éléments est nécessaire, les deux points remplissant le même rôle d'explication que la conjonction de coordination. Quel élément le transcripteur doit-il garder ? La ponctuation initiale de l'élève, sachant que maîtriser les deux points à cet âge est plutôt rare ? Ou bien la conjonction de coordination ?

Une autre difficulté dans ce texte tient à l'interprétation que l'on peut faire du segment *Mathilde et ses amis sortirent de la maison quand les parents de Mathilde leur demandèrent où est-ce qu'ils allaient il avait une poupée qui était par terre* (lignes 14 à 17 de la copie). Si

l'on considère que les parents de Mathilde demandent aux enfants leur destination au moment où ils sortent de la maison, le segment sera découpé comme suit :

> *Mathilde et ses amis sortirent de la maison quand les parents de Mathilde leur demandèrent où est-ce qu'ils allaient. Il avait une poupée qui était par terre.*

Si l'on considère qu'au moment où les parents de Matilde posent la question, les enfants voient une poupée par terre, le segment sera découpé de la manière suivante :

> *Mathilde et ses amis sortirent de la maison. Quand les parents de Mathilde leur demandèrent où est-ce qu'ils allaient, il avait une poupée qui était par terre.*

On le voit bien, toute transcription entraîne des choix. Nous proposons donc pour ce texte la transcription normée suivante :

Elle habitait dans cette maison depuis longtemps. Mathilde avait déménagé de Tournefeuille pour aller habiter dans une maison sinistre. Quelques jours plus tard, elle avait mis toutes ses affaires ainsi que sa famille. Mathilde n'aimait pas dormir dans sa chambre qu'elle avait décorée car il y avait grincement qui faisait froid dans le dos. Elle appela ses nouveaux amis pour avoir de la compagnie. Quand soudain il entendit un bruit. *Il se retourna en entendant ce grand bruit.* Mathilde et ses amis sortirent de la maison quand les parents de Mathilde leur demandèrent où est-ce qu'ils allaient. Il y avait une poupée qui était par terre. Ils la ramassèrent. La poupée avait un livre en main. Elle était déguisée en sorcière. Elle sortit. Elle avait les yeux rouges. Ils déchirèrent la poupée et rentrèrent chez eux. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 24 : Transcription normée, Texte de Rayan, CM1 T2

Le rétablissement d'une ponctuation normée dans les textes d'élèves, outre les difficultés qu'elle soulève pour le linguiste, montre que la notion de segmentation des phrases n'est pas simple pour le scripteur en cours d'apprentissage. C'est une acquisition qui se fait progressivement sur les différents niveaux scolaires (Jaffré & Fayol, 1989 ; Cappeau & Roubaud, 2005 ; Garcia-Debanc, 2010).

Au terme de ce chapitre, nous avons présenté nos questions de recherche ainsi que notre protocole de recueil de données. Nous avons également détaillé les étapes de transcription du corpus d'étude, notamment à travers la double transcription effectuée (diplomatique et normée). Dans le chapitre suivant, nous allons exposer les critères de notre analyse.

CHAPITRE 7 : Critères pour l'analyse de la référence dans les textes d'élèves et problèmes d'interprétation

Dans notre recherche, nous nous proposons d'étudier comment les élèves introduisent et maintiennent les référents dans les productions écrites narratives. Nous étudions les expressions référentielles, c'est-à-dire nous nous appliquons à « identifier par quel mécanisme une entité du discours est désignée par une mention linguistique » (Landragin, 2014, s'appuyant sur Charolles, 2002). Lorsque des mentions linguistiques sont relatives à la même entité du discours, elles forment une chaîne de coréférence, appelée *anaphore* si la chaîne contient 2 maillons ou *chaîne de référence* (CDR) si elle en contient au moins 3 (cf. chapitre 2). Notre recherche propose donc d'étudier à la fois la première mention d'un référent (ou comment le référent est introduit) et la succession des mentions du référent à l'intérieur d'une chaîne de coréférence (ou comment le référent est maintenu dans le discours).

Afin d'analyser les expressions référentielles dans notre corpus, nous avons choisi d'annoter manuellement chaque texte à l'aide du logiciel de tableur Excel, sous forme de tableau et selon une grille de catégories que nous nous apprêtons à décrire dans ce chapitre.

De prime abord, le choix du logiciel de tableur Excel peut paraître désuet car, comme l'affirment Mélanie-Bécquet & Landragin (2014 : 123) : « si c'est un moyen comme un autre pour appliquer sur un texte des annotations […] ce n'est évidemment pas le moyen le plus approprié et le plus pérenne ». Pourtant, au vu des spécificités de notre corpus de textes d'élèves (cf. chapitre 5, section 5.2.) et de la consigne précise qui restreint fortement notre champ de recherche, Excel permet le degré d'annotation dont nous avons besoin et toutes les analyses statistiques que la plupart des logiciels d'annotation proposent. Il faut simplement s'adapter à son ergonomie qui n'a pas été initialement prévue pour ce type de tâche et repenser la logique de présentation en créant des codes caractéristiques correspondant à chacune des catégories étudiées.

En outre, l'annotation étant un phénomène « empirique » ou « cyclique » (Landragin, 2011), il est nécessaire d'adapter les catégories d'analyse au fur et à mesure du travail, ce que permet facilement ce type de logiciel. Enfin, comme l'explique Landragin (2014), l'étude des phénomènes complexes de référence et chaîne de coréférence ne se fait que rarement avec l'exploitation d'un logiciel unique de type outil d'aide à l'annotation, mais bien dans l'aller-

retour entre un ou plusieurs de ces outils et des logiciels de type tableur tel qu'Excel. À ce sujet, nous renvoyons à l'étude spécifique des chaînes de coréférence à l'aide des outils combinés GLOZZ, ANALEC et Excel, *in* Landragin, 2014.

Malgré les difficultés propres à l'analyse de la référence et de la coréférence que nous venons de décrire, pour réaliser une étude sur ce sujet, nous avons analysé particulièrement un certain nombre d'informations. Aussi en 7.1., nous décrivons les différentes catégories d'analyse utilisées pour l'annotation de notre corpus. Malgré tout, nous verrons en 7.2. que des difficultés demeurent dans l'attribution des référents et posent des problèmes d'interprétation.

7.1. Catégories d'annotation

Dans la tâche problème de notre étude, les référents à introduire sont présents dans les phrases de la consigne. Il s'agit des référents des pronoms personnels *elle* et *il*, du référent du SN défini *les enfants*, des référents des SN démonstratifs *cette maison* et *ce grand bruit* et du référent de l'indication temporelle *la nuit*. Les référents des entités de la consigne sont désignés à l'aide d'expressions référentielles. Pour rappel, une expression référentielle est « un ensemble de mots consécutifs du texte, qui permet d'identifier et donne accès à un référent, et permet ainsi à la phrase ou à la proposition courante de dire quelque chose de ce référent. » (Landragin & *al.* 2015 ; cf. aussi chapitre 1, section 1.4.).

Pour réaliser notre étude de la référence dans les textes d'élèves, nous avons mis en place divers critères d'annotation inspirés de Schnedecker & Longo (2012), que nous avons regroupés sous la forme d'un tableau via le logiciel de tableau Excel. Nous utilisons un tableau d'analyse par occurrence à traiter. Cette présentation nous permet d'avoir toutes les données d'une occurrence en une seule vue. Chaque colonne de nos tableaux correspond à une catégorie d'analyse. Dans cette section, nous proposons d'exposer chacune de ces catégories.

En 7.1.1., nous détaillons les catégories générales qui regroupent les métadonnées et les éléments propres à l'ensemble d'un texte. En 7.1.2., nous exposons les catégories qui servent à l'annotation de la nature des formes linguistiques de chacune des mentions du référent. En 7.1.3., nous présentons les catégories d'annotation employées pour l'analyse des

expansions des expressions référentielles. Enfin, en 7.1.4., nous exposons les catégories d'analyse des fonctions syntaxiques de chacune des mentions du référent.

7.1.1. Informations générales à l'échelle du texte entier

Pour chaque texte de notre corpus, nous regroupons un certain nombre d'informations générales afin qu'il puisse être analysé de façon autonome. D'une part, des métadonnées concernant le rédacteur et d'autre part, des informations propres à notre analyse.

En ce qui concerne les métadonnées, nous précisons le prénom seul de l'élève, son sexe, sa classe et le niveau scolaire (CE2, CM1 ou CM2).

Pour chaque texte, nous indiquons dans notre tableau diverses informations générales :

- Si le texte est exploité ou non : les textes non exploités peuvent être des textes dans lesquels la consigne n'a pas été respectée (par exemple en changeant l'ordre des phrases ou en les collant à la suite les unes des autres), ou bien des textes illisibles contenant un trop grand nombre d'éléments indéchiffrables.
- Le nombre total de mots (comptage dans la version normée) : le comptage du nombre de mots permet de pouvoir comparer les longueurs des textes entre plusieurs niveaux de classe ou entre des élèves d'un même niveau.
- Si le texte est achevé ou non
- Si la phrase imposée P1 est la première phrase du texte de l'élève

Nous indiquons également la localisation de la source du référent, c'est-à-dire à quel endroit du texte se trouve la première mention du référent : avant la phrase imposée, après la phrase imposée, ou bien si le texte ne contient aucune autre mention du référent que celle de la phrase imposée.

7.1.2. Forme des mentions du référent

Pour chaque référent de la consigne, nous annotons chacune de ses mentions tout au long du discours et notamment, sa forme linguistique. De cette manière, nous nous appuyons sur Charolles (2002) pour déterminer, pour chacune des mentions, s'il s'agit d'un :

- SN indéfini : *un/une/des* + N
- SN défini : *le/la/les* + N
- SN possessif : *son/sa/ses* + N
- SN démonstratif : *ce*/*cet/cette* + N
- SN pronominal : *elle/il/le/la/leur/…*
- Nom propre

De surcroît, les personnages de la consigne peuvent s'associer dans le récit et former des groupes. Par exemple, le référent du SN défini *les enfants* peut être le groupe constitué par les référents des pronoms personnels *elle* et *il.* Le groupe ainsi constitué peut alors être la première entité désignée dans le texte. Ainsi, lorsqu'il est question de référence à un groupe de personnages, nous faisons le choix de la considérer comme une mention du référent car selon Landragin (2011 : 68) : « les pluriels et les groupes d'individus impliqués construisent des entités du discours au même titre qu'une expression référant à un individu unique ». Pour cette raison, nous choisissons d'annoter également les groupes pluriels que nous rencontrons, sous la catégorie *SN numéral*.

Pour l'occurrence *ce grand bruit*, le sémantisme du nom *bruit* peut entraîner l'emploi d'onomatopées. Nous les annotons sous la catégorie éponyme.

A ces catégories, il est vrai que nous pourrions ajouter les marques d'accord en genre et en nombre car elles participent à la référence en rappelant la nature du référent sujet. Dans *ils dorment*, la terminaison *–ent* rappelle que le sujet du verbe est au pluriel, et par là-même rappelle le référent *ils*. Toutefois, à l'instar de Landragin (2011), nous choisissons de les ignorer « compte-tenu de l'aspect grammaticalisé et nécessaire de ces phénomènes ». Il en va de même pour les constructions pronominales telles que *se laver*, *se promener*… considérées comme faisant partie du verbe.

En revanche, nous prenons en compte les expressions telles que les sujets zéro. Les sujets zéro peuvent être saillants et contribuer fortement aux relations de coréférence. À ce titre, ils

peuvent être considérés comme élément référentiel linguistiquement non marqué (Landragin, 2011) car le référent reste saillant malgré l'absence de marque linguistique.

En outre, notamment pour les entités non humaines et non animées, à savoir *cette maison*, *ce grand bruit* et *la nuit,* nous pouvons rencontrer des cas d'inférence. Il s'agit de textes dans lesquels aucune marque linguistique ne mentionne le référent mais où le lecteur comprend tout de même la référence à l'entité visée. Nous choisissons de regrouper ces cas sous la catégorie *Autre* et d'apporter des précisions sur le type d'inférence à effectuer :

- inférence portant sur le verbe : *habiter* par exemple pour l'occurrence *maison*
- inférence portant sur le nom : *étoiles* par exemple pour l'occurrence *la nuit*
- phrase entière explicative : *Il entendit les pommes tomber* par exemple pour l'occurrence *grand bruit*.

Par surcroît, pour l'occurrence *ce grand bruit*, nous précisons s'il y a la présence du verbe *entendre* ou non. En effet, le degré d'inférence - et par là même, le degré de maîtrise des procédés référentiels - n'est pas le même si le calcul interprétatif s'appuie sur un nom seul, sur un nom + un verbe ou sur un nom + le verbe *entendre*.

À l'instar de l'étude de Landragin & *al.* (2015), nous faisons donc le choix de prendre en compte les formes explicites de la référence mais aussi les formes atténuées, ces deux types de formes « constituant les deux catégories de *maillons* qui vont […] servir à construire puis à étudier les chaînes de coréférence » (*Ibid.* : 3).

Pour résumer, dans la catégorie des formes des mentions du référent, nous trouvons :

- SN indéfini : *un/une/des* + N
- SN défini : *le/la/les* + N
- SN possessif : *son/sa/ses* + N
- SN démonstratif : *ce*/*cet/cette* + N
- SN pronominal : *elle/il/le/la/leur/…*
- Nom propre
- SN numéral : *nombre* + N
- Onomatopées
- Autre : divers cas d'inférence

À côté de chacune des formes linguistiques, nous inscrivons la mention telle qu'elle est écrite par l'élève afin de pouvoir réaliser par la suite une étude des formes lexicales.

7.1.3. Modifieurs des mentions du référent

Lorsqu'on étudie la référence se pose aussi le problème de la délimitation d'une expression référentielle. En effet, annoter des phénomènes référentiels c'est délimiter les maillons des chaînes de référence ou coréférence pour ensuite leur attribuer des propriétés. (Mélanie-Becquet & Landragin, 2014). La question de la délimitation des expressions référentielles est importante car elle conditionne ce qui relève de la mention au personnage (ou à l'entité dont on parle) et ce qui relève d'une expansion de cette mention (Landragin 2011), dans le but d'apporter des précisions sur l'objet du discours.

Dans le cadre de notre tâche-problème, toutes les expansions n'ont pas le même « poids ». En effet, le fait de savoir que le personnage féminin est plutôt gentil ou méchant, ou bien qu'il déménage par exemple, n'influe pas sur le sens du pronom personnel *elle* de P1. En revanche, une expansion de la mention qui réfère au SN démonstratif *ce grand bruit* de P2 a toute son importance car c'est elle qui va déterminer si le lecteur sera d'accord pour dire que le bruit introduit correspond bien à un *grand* bruit et non pas à un bruit que l'on pourrait qualifier de faible.

Aussi, lorsqu'une mention du référent est expansée, nous l'annotons selon une typologie qui rend compte de sa catégorie grammaticale :

- Syntagme adjectival
- Syntagme prépositionnel
- Proposition relative
- Nom propre

Si la mention du référent comporte plusieurs modifieurs, nous annotons chacun d'entre eux séparément : *forme de l'expansion 1*, *forme de l'expansion 2*, etc.

À côté de chacune des expansions, nous inscrivons la mention telle qu'elle est écrite par l'élève afin de pouvoir réaliser par la suite une étude lexicale des expansions employées.

7.1.4. Fonctions syntaxiques des mentions du référent

Nous l'avons vu dans le chapitre 1, la position syntaxique d'une entité est en lien direct avec le statut informationnel. En effet, un principe universel d'organisation des langues (Lambrecht, 1994 ; Hickmann, 1995 ; Jisa, 2000 ; Khorounjaia & Tolchinsky, 2004) veut que l'information ancienne soit située préférentiellement en début de phrase, notamment en position de sujet syntaxique et très souvent sous la forme d'un pronom (Chafe, 1994). L'information nouvelle serait, elle, plutôt en fin de proposition, en position non sujet, car elle nécessite davantage de ressources cognitives. Du point de vue de la progression thématique, le thème peut correspondre au sujet grammatical, au niveau des référents (Lyons, 1980) et doit être en position initiale dans la proposition (Halliday, 1967). Le rhème correspond ordinairement au prédicat. Enfin, la fonction grammaticale d'un mot ou groupe de mots influe aussi sur son degré de saillance. Dans une langue à structure SVO (Sujet-Verbe-Objet) comme le français, le sujet syntaxique, ordinairement en première position dans la phrase, est considéré comme le constituant grammatical le plus saillant. La Théorie du Centrage d'attention (Grosz *& al.*, 1995) suggère d'ailleurs qu'après le sujet, le complément d'objet est l'élément le plus saillant, suivi par les autres catégories grammaticales.

Dans le chapitre 3, nous avons montré que ce n'est qu'à partir de l'âge de 6 ans que les enfants commencent à utiliser des pronoms anaphoriques pour maintenir la référence, à travers la stratégie du sujet thématique (Karmiloff-Smith, 1980 ; 1981). En effet, les enfants de 6 ans et au-delà réservent l'emploi du pronom personnel placé en début de phrase au sujet thématique, qui est en général le personnage principal de leur histoire. Les autres entités du récit sont dénommées à l'aide de SN définis et placées dans des positions syntaxiques diverses.

Dans les phrases de notre tâche-problème, les référents que les élèves doivent introduire sont placés dans des positions syntaxiques « ordinaires » : les pronoms personnels *elle* et *il* ainsi que le SN défini *les enfants* sont en position de sujet syntaxique, les SN démonstratifs *cette maison* et *ce grand bruit* sont en position de complément de verbe et le SN défini *la nuit* en position de complément de phrase. Nous retrouvons donc ces différentes fonctions syntaxiques dans nos catégories d'analyse.

Aux catégories de fonctions syntaxiques que nous venons d'évoquer, nous ajoutons la *construction présentative* (Lambrecht, 2000). Il s'agit d'une construction dont la fonction est

de « présenter une entité nouvelle dans un discours donné et d'exprimer une information nouvelle au sujet de cette entité » (*ibid.* : 49). L'ajout de cette catégorie dans les fonctions syntaxiques s'explique par deux raisons. La première tient au fait que la consigne demande aux élèves de *raconter* une histoire. Parmi les différents types de littérature auxquels les élèves sont acculturés, beaucoup d'histoires sont des contes. Les élèves vont donc avoir tendance à en reprendre les codes. Cela se traduit par des textes dans lesquels la première phrase débute souvent par la construction présentative *Il était une fois*. La deuxième raison tient au fait que dans les textes d'apprentis scripteurs, on trouve souvent des tournures de phrase typiques de l'oral (Blanche-Benvéniste, 1997), telles que les constructions clivées en *c'est X qui*, ou *c'était* ou encore les constructions en *il y a/avait*. Ce type de constructions introduit une entité ou une situation nouvelle dans le but de la rendre cognitivement accessible en vue d'une prédication ultérieure (Lambrecht, 2000). Ce sont donc également des constructions présentatives.

Enfin, nous ajoutons à ces catégories, d'une part, les *phrases nominales* et d'autre part, la fonction *complément du nom.*

Ainsi, l'étude de la référence dans les textes de notre corpus nous conduit à annoter la fonction syntaxique de chaque mention du référent. Nous recensons donc les catégories suivantes :

- Sujet
- Construction Présentative
- Complément de verbe
- Complément de phrase
- Complément du nom
- Phrase nominale

7.2. Problèmes d'interprétation référentielle

Dans l'élaboration des critères d'annotation définis *supra*, certains cas résistent et il est difficile de les faire correspondre à une catégorie d'analyse. En effet, comme nous l'avons déjà explicité (cf. chapitre 5), les textes d'apprentis scripteurs de notre corpus soulèvent beaucoup de difficultés, notamment à cause de l'aspect non normé des productions. Aux

niveaux sémantique et référentiel, les difficultés soulevées engendrent des problèmes d'interprétation parfois complexes.

Dans notre recherche, nous annotons chaque mention des référents de la consigne. En d'autres termes, nous repérons les expressions référentielles qui désignent les référents des phrases imposées et nous les analysons. Cependant, il est parfois difficile de voir « à quel référent introduit ou repris dans le discours précédent ou suivant peut renvoyer une expression anaphorique » (Charolles, 2014 : 55). Plusieurs raisons expliquent ces difficultés. Nous proposons d'en rendre compte dans cette section, en les illustrant par des extraits de textes d'élèves de notre corpus.

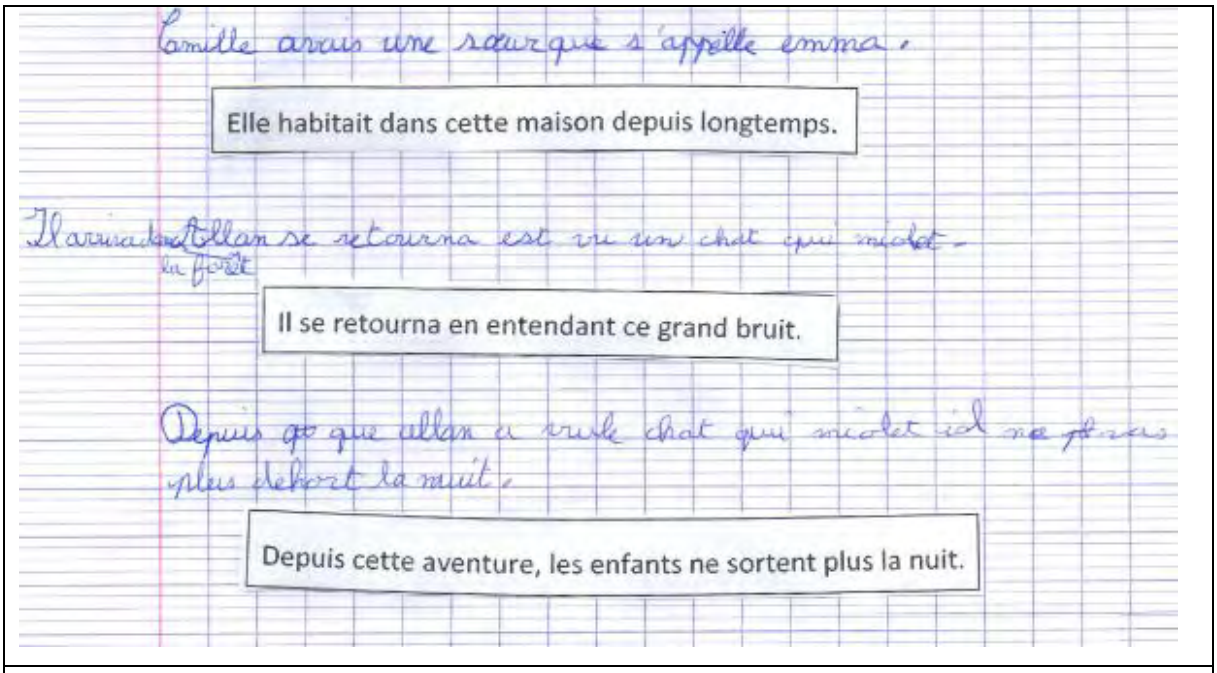
En 7.2.1., nous présentons les cas d'ambiguïté référentielle dus à la présence de plusieurs candidats à la référence. En 7.2.2., nous montrons comment le non-respect des contraintes morphosyntaxiques imposées par la consigne entraîne des difficultés d'interprétation référentielle. En 7.2.3., nous nous intéressons à la question des contradictions entre les marques syntaxiques du genre et la référence attendue. En 7.2.4., nous exposons les problèmes d'accessibilité des référents causés par le choix des expressions référentielles. En 7.2.5., nous indiquons comment un calcul interprétatif trop coûteux peut faire échouer la référence. Enfin, en 7.2.6., nous évoquons la discordance sémantique et les problèmes qu'elle engendre dans l'attribution d'un référent.

7.2.1. Ambiguïté référentielle : plusieurs candidats à la référence

Lors de la lecture d'un texte, généralement, c'est au moment où l'interlocuteur rencontre l'expression anaphorique qu'il cherche dans sa mémoire discursive la bonne source. Il ne revient pas en arrière dans le texte (ou le discours) pour y trouver l'endroit exact où est localisé la source (cf. chapitre 2, section 2.1.1.2.). Toutefois, ce retour en arrière s'effectue tout de même lorsque la résolution de l'expression anaphorique est problématique, comme dans les exemples ci-dessous.

Dans (25), l'ambiguïté tient au fait qu'il existe deux candidats possibles à la référence du pronom personnel *elle* de P1 : les noms propres *Camille* et *Emma.* Si l'on part du principe que le nom propre *Camille* est un nom masculin, le problème ne se pose pas. Mais puisque c'est un prénom mixte, on ne peut le rejeter trop vite. Les deux noms propres employés dans

la première phrase sont donc tous deux de genre féminin et sont ainsi autant susceptibles l'un que l'autre d'être le référent du pronom personnel *elle.*

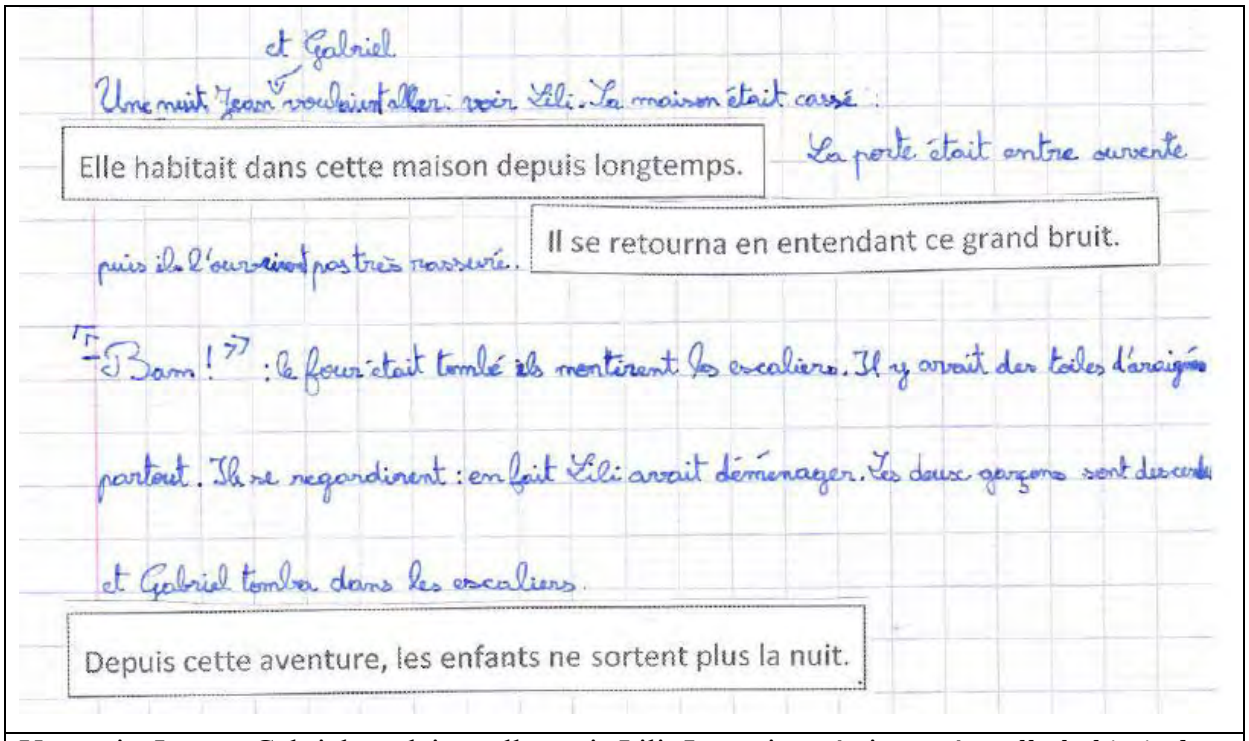


Camille avait une sœur qui s'appelle Emma. *Elle habitait dans cette maison depuis longtemps.* Il arrivait dans la forêt. Allan se retourna et vit un chat qui miaulait. *Il se retourna en entendant ce grand bruit.* Depuis qu'Allan a vu le chat qui miaulait, il ne va plus dehors la nuit. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 25 : Texte de Romane, CE2 L

Pour tenter de comprendre à quel référent le pronom fait appel, il peut y avoir deux manières de procéder. Lorsqu'on examine la progression thématique (cf. chapitre 1, section 1.2.3.), dans le premier énoncé, on voit que *Camille* est en position initiale dans l'énoncé, c'est donc le thème. Le rhème du premier énoncé, *Emma*, devient le thème du second énoncé (P1). On observe donc une progression linéaire. Dans cette première interprétation, c'est le nom propre *Emma* qui est le référent du pronom personnel *elle.* Il s'agit de la stratégie de distance minimale (Kail & Léveillé, 1977 ; cf. chapitre 3, section 3.1.1.). Mais si on se place dans la perspective d'une progression à thème constant, le thème du premier énoncé, *Camille*, reste le thème de l'énoncé suivant (P1). Dans cette seconde interprétation, c'est le nom propre *Camille* qui est le référent du pronom personnel *elle*. Il s'agit alors de la stratégie de non-changement de rôle (Kail & Léveillé ; 1977, cf. chapitre 3, partie 3.1.1.).

De la même manière, dans (26), l'ambiguïté référentielle tient au fait qu'il y a deux candidats à la référence :



Une nuit, Jean et Gabriel voulaient aller voir Lili. La maison était cassée. *Elle habitait dans cette maison depuis longtemps.* La porte était entrouverte puis ils l'ouvrirent, pas très rassurés. *Il se retourna en entendant ce grand bruit.* « Boum ! » : le four était tombé. Ils montèrent les escaliers. Il y avait des toiles d'araignées partout. Ils se regardèrent : en fait Lili avait déménagé. Les deux garçons sont descendus et Gabriel tomba dans les escaliers. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 26 : Texte d'Estelle, CM1 T2

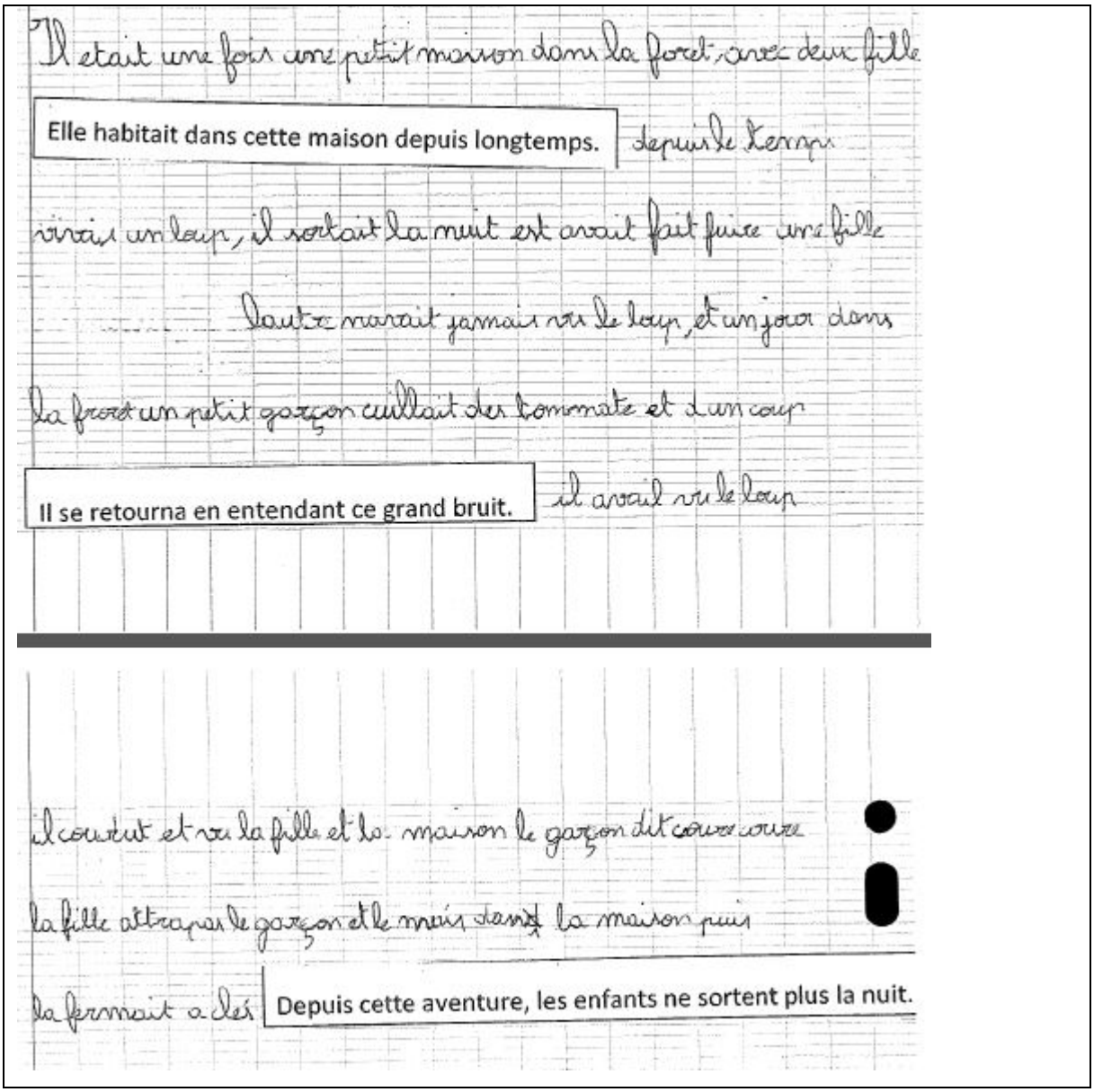
Dans la première phrase l'élève introduit deux personnages masculins, *Jean* et *Gabriel*. Dans la phrase suivante, l'élève réfère au groupe que forment ces deux personnages, par l'emploi du pronom personnel *ils*. Mais quand vient P2, l'élève n'a pas réinstancié l'un ou l'autre de ses personnages masculins. Le lecteur ne peut donc savoir qui, de *Gabriel* ou de *Jean* a entendu le bruit.

7.2.2. Non perception des contraintes morphosyntaxiques

L'exemple (26) ci-dessus soulève également un autre problème. En effet, dans la suite du texte, il est seulement question des référents au pluriel, désignés par le pronom personnel *ils*. Mis à part dans P2, il n'est pas question d'un référent masculin singulier ailleurs dans le texte. Il semble donc qu'il y a non perception de la marque du singulier dans P2, portée non

seulement par le pronom personnel singulier *il* mais aussi par la terminaison du verbe *se retourner* dans *il se retourna*. On remarque en revanche que la réinstanciation s'opère pour la chute avec le nom propre *Gabriel*.

Le même cas se retrouve dans l'exemple (27) au sujet du référent féminin de P1.


Il etait une fois une petit maison dans la foret, avec deux fille

Elle habitait dans cette maison depuis longtemps.

depuis le temps

vivait un loup, il sortait la nuit est avait fait fuire une fille

l'autre navait jamais vu le loup, et un jour dans

la foret un petit garçon cueillait des tomate et d'un coup

Il se retourna en entendant ce grand bruit.

il avait vu le loup

il courut et vu la fille et la maison le garçon dit cours cours

la fille attrapa le garçon et le mis dans la maison puis

la fermait a clés

Depuis cette aventure, les enfants ne sortent plus la nuit.

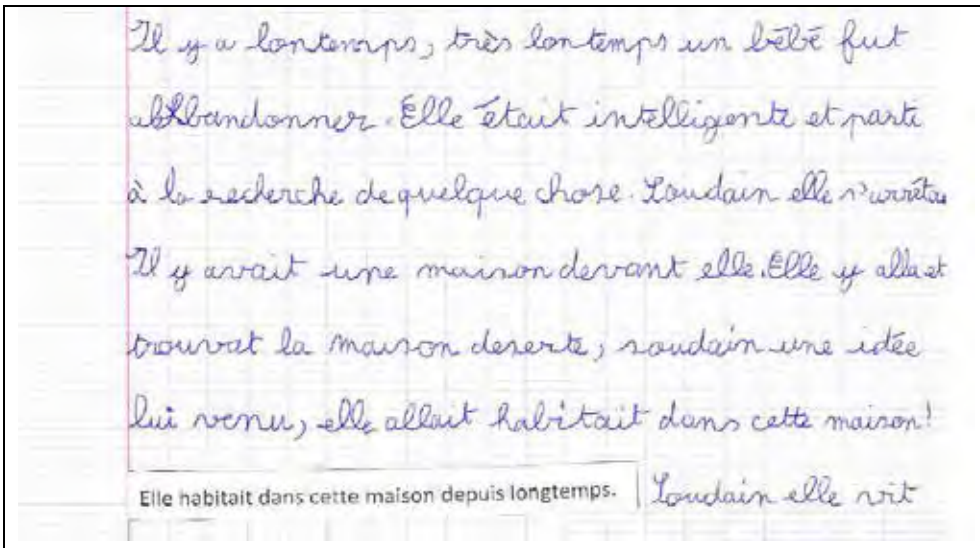
Il était une fois une petite maison dans la forêt avec deux filles. *Elle habitait dans cette maison depuis longtemps.* Depuis le temps vivait un loup, il sortait la nuit et avait fait fuir une fille. L'autre n'avait jamais vu le loup et un jour, dans la forêt, un petit garçon cueillait des tomates et d'un coup *il se retourna en entendant ce grand bruit.* Il avait vu le loup. Il courut et vit la fille et la maison. Le garçon dit : « Cours ! Cours ! ». La fille attrapa le garçon et le mit dans la maison puis ferma à clé. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 27 : Texte de Soliana, CM1 T1

Ici, le seul candidat à la référence pour le pronom personnel *elle* de P1 est le SN numéral *deux filles* dans la phrase précédent P1. Mais il est au pluriel alors que le pronom personnel est au singulier. On peut penser de prime abord qu'il y a deux candidats à la référence mais que l'élève a oublié de les distinguer. Pourtant, dans la suite du texte, les deux personnages féminins sont différenciés : *une fille*/*l'autre*. Il ne s'agit donc pas d'un problème référentiel impliquant plusieurs candidats mais d'un problème de non prise en compte du genre imposé par la consigne.

7.2.3. Contradiction entre les marques syntaxiques du genre et la référence attendue

Dans certains textes, la différence de genre entre le référent et une (ou plusieurs) de ses mentions n'est pas due à la non perception des marques morphosyntaxiques mais à un choix de la part du scripteur. Lorsqu'il est mal contrôlé, ce choix peut entraîner des problèmes d'interprétation. Prenons un texte comme (28) :


Il y a lontemps, très lontemps un bébé fut
abbbandonner. Elle était intelligente et parti
à la recherche de quelque chose. Soudain elle s'arrêta
Il y avait une maison devant elle. Elle y alla et
trouvat la maison deserte, soudain une idée
lui venu, elle allait habitait dans cette maison!
Elle habitait dans cette maison depuis longtemps. | Soudain elle vit

Il y a longtemps, très longtemps, un bébé fut abandonné. Elle était intelligente et partit à la recherche de quelque chose. Soudain, elle s'arrêta. Il y avait une maison devant elle. Elle y alla et trouva la maison déserte. Soudain, une idée lui vint : elle allait habiter dans cette maison ! *Elle habitait dans cette maison depuis longtemps.*

Figure 28 : Texte de Romain, CE2 P

Dans cet exemple, le premier personnage est introduit sous la forme du SN indéfini masculin *un bébé* dans la première phrase. Dans la deuxième phrase, le pronom personnel féminin *elle* apparaît. À cet endroit du texte, on peut donc se poser la question de savoir si le SN *un bébé* est un bon candidat pour le pronom anaphorique. Mais lorsqu'on lit le texte jusqu'à P1, on comprend que le pronom personnel *elle* employé dès la deuxième phrase réfère bien au même personnage que le pronom personnel *elle* de P1, puisqu'il n'y a aucun autre candidat à la référence. Le pronom étant une forme de haute accessibilité, il est nécessaire de l'introduire au préalable par une expression coréférentielle de moindre accessibilité, telle qu'un SN indéfini par exemple. Dans le texte, le seul candidat possible est le SN indéfini *un bébé*, mais nous avons vu qu'il y avait un problème de genre.

Pourtant, dans le texte, on comprend bien que le référent du SN indéfini *un bébé* est le même que celui du pronom personnel *elle* des phrases suivantes. Cette interprétation est accentuée par le sémantisme du nom *bébé.* D'une part, malgré le déterminant masculin devant le mot *bébé*, dans le monde réel, l'entité dénotée par le mot *bébé* peut être de sexe féminin ou bien de sexe masculin. De ce fait, dans le langage, il est courant d'employer un pronom personnel masculin ou féminin pour référer au SN *bébé*. D'autre part, le SN *bébé* dénote l'état d'une entité plus que l'entité elle-même : on sait qu'un bébé est un être humain à un certain stade de sa croissance et que cet être humain est appelé à évoluer. Toutefois, si l'on s'en tient

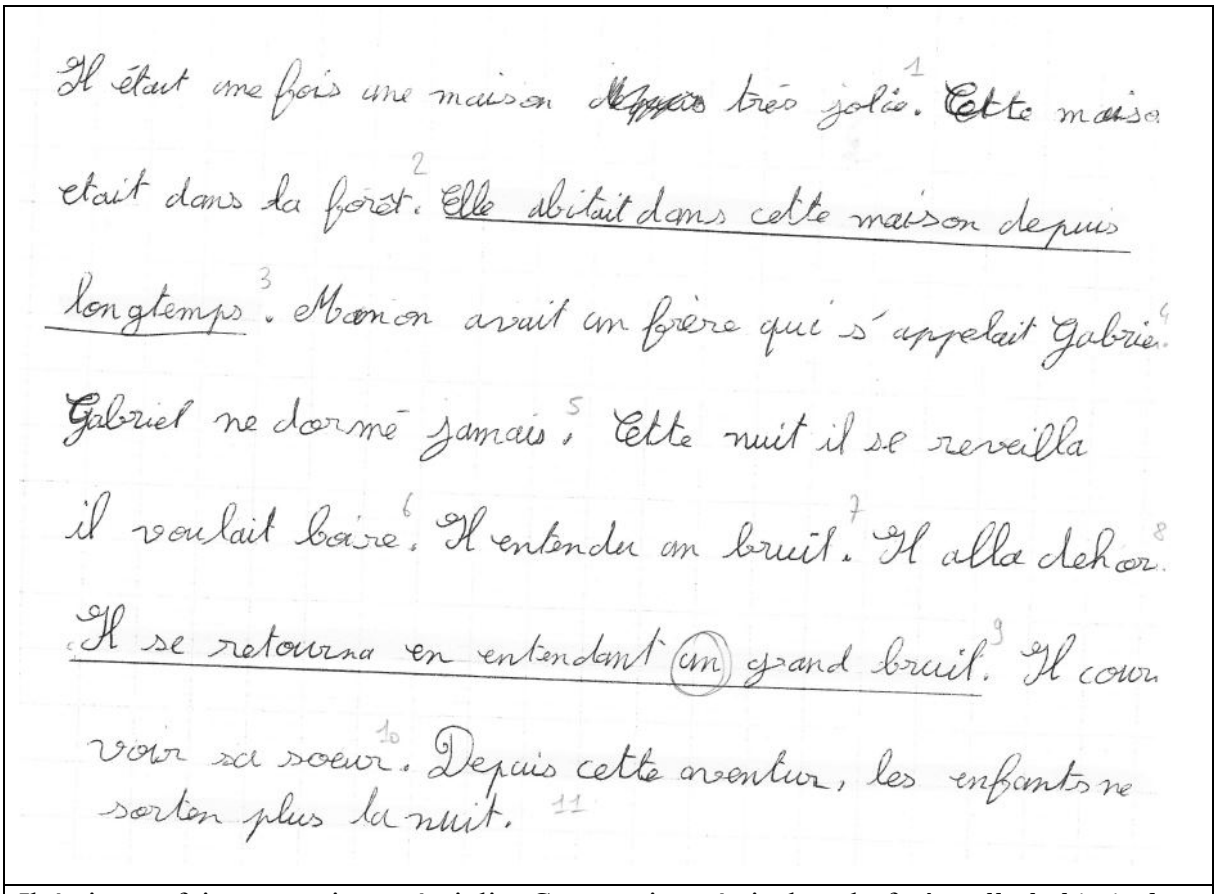
à une question d'identité référentielle, on ne peut nier que le SN indéfini *un bébé* est différent de l'entité à laquelle réfère le pronom personnel *elle* : le référent du pronom personnel est un autonome dans le déplacement et capable d'intention (*elle était intelligente et partit*), ce ne peut donc être un *bébé*. De la même manière, Brown & Yule (1983 ; cf. chapitre 2, section 2.1.1.1) se demandaient si le poulet « actif et bien gras » de départ d'une recette de cuisine était bien le même poulet rôti à l'arrivée, malgré les diverses transformations qu'il avait subies.

Pour autant, on ne peut parler de référent évolutif dans ce cas (Yule, 1982 ; Brown & Yule, 1983 ; Schnedecker & Charolles, 1993 ; Charolles, 2001) car le verbe de la phrase hôte du référent, *abandonné*, ne comporte pas de prédicat transformateur visant le référent. Malgré l'absence de prédication transformatrice, le changement de genre entre le SN indéfini *un bébé* et le pronom personnel *elle* peut être résolu par un calcul inférentiel de la part du lecteur. L'attribution du référent *bébé* à l'expression anaphorique *elle* passe par la représentation mentale que s'en fait l'interlocuteur, et non plus seulement par le discours préalable (Reboul, 1989 ; cf. chapitre 2, section 2.1.1.2.).

7.2.4. Problème d'accessibilité du référent dans le choix des formes linguistiques des mentions

Dans certains cas, les textes différencient bien les référents dont il est question mais il reste difficile de comprendre du premier coup à quel référent du discours peut renvoyer l'expression anaphorique.

C'est le cas par exemple dans (29), où le seul référent possible avant P1 pour interpréter le pronom personnel *elle* de P1 est le SN *une maison* car il est du même genre et du même nombre. Le lecteur s'attend donc à ce que ce soit le sujet grammatical des phrases suivantes. Or, on comprend bien que la *maison* dont il est question dans les deux premières phrases ne peut pas être le référent du pronom personnel *elle* de P1, puisqu'une maison ne peut pas habiter dans une maison. Le seul candidat possible à la référence, une entité humaine de même genre et de même nombre, se trouve après P1, sous la forme du nom propre *Manon.* On est ici dans le cas d'une introduction cataphorique (cf. chapitre 2, section 2.1.).



Il était une fois une maison très jolie. Cette maison était dans la forêt. *Elle habitait dans cette maison depuis longtemps.* Manon avait un frère qui s'appelait Gabriel. Gabriel ne dormait jamais. Cette nuit, il se réveilla, il voulait boire. Il entendit un bruit. Il alla dehors. *Il se retourna en entendant ce grand bruit.* Il court voir sa sœur. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 29 : Texte de Chemsa, CE2 U

La difficulté d'interprétation ici ne tient pas tant au fait que le référent est introduit après P1, mais plutôt au contexte de l'introduction. En effet, comme indiqué précédemment, la saillance du SN *maison* induite par la construction présentative de la première phrase fait que l'on s'attend à ce que cette entité soit le sujet des propositions qui suivent. Lorsque le lecteur arrive à P1, il est décontenancé par la présence du pronom personne *elle* en position saillante et il échoue à construire la référence. Mais lorsqu'il continue la lecture et qu'il rencontre le nom propre *Manon*, qui est le bon référent, il peut également être décontenancé par le manque de cohérence entre P1 et la phrase qui suit. En effet, un nom propre s'emploie plutôt pour introduire ou réinstancier un personnage. Or, dans la phrase précédente (P1), le personnage féminin est mentionné à l'aide d'un pronom personnel, qui est une forme employée lorsque le degré d'accessibilité du référent est la plus forte (cf. degré d'accessibilité

des référents, chapitre 1, section 1.2.2). On passe donc d'une phrase où le personnage féminin est présenté comme une information connue, accessible (pronom personnel *elle*) à une phrase dans laquelle il est présenté comme moins accessible qu'auparavant (nom propre *Manon*), alors que cela devrait être l'inverse. L'échec de la construction de la référence dès la première lecture constitue un facteur de manque de cohérence textuelle, que le lecteur indulgent peut combler par un calcul interprétatif.

7.2.5. Inférence avec calcul interprétatif coûteux

Le point précédent nous amène à une question soulevée lors de l'annotation de notre corpus : quel degré de coopération peut-on attendre de la part du lecteur lorsqu'il s'agit d'effectuer un calcul interprétatif ?

Dans (30) par exemple, à la première lecture, le lecteur peut décréter qu'il n'y a pas de référent au pronom personnel *elle*.

Elle habitait dans cette maison depuis longtemps. Il y eut du bruit, personne ne l'entendit sauf un : c'était Michel. *Il se retourna en entendant ce grand bruit.* Ce vacarme venait de sa voisine, elle était tombée dans les escaliers avec un fil. Elle s'est fracassée la tête ouverte et c'était les deux chenapans, c'était Romain et Thomas qui sont allés en prison pendant 15 jours. *Depuis cette aventure, les enfants ne sortent plus la nuit.* Pour ne plus faire de bêtise et ne pas aller en prison. FIN

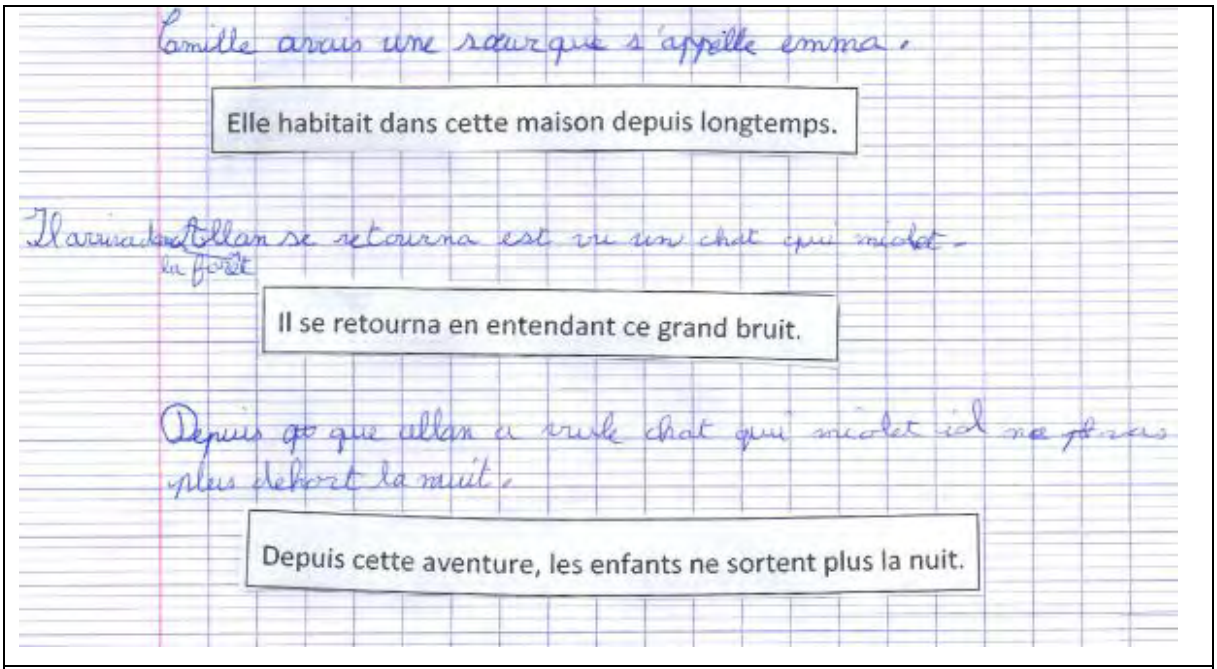
Figure 30 : Texte de Suzanne, CM1 B

Pourtant, si l'on relit attentivement le texte, on peut être tenté de reconstituer une continuité référentielle entre le pronom personnel *elle* et le SN *voisine.* Dans l'histoire, Michel entend un bruit qui vient de sa voisine, qu'il indique être tombée dans l'escalier. S'il a entendu le bruit, on peut inférer qu'elle habite une maison à côté de la sienne. Or, même si P1 semble éloigné de la première mention du personnage de la voisine, le SN *maison* de P1 et le sémantisme du mot *voisine*, d'autant que c'est le seul candidat à une possible référence pour *elle*, peuvent amener à identifier le personnage de la voisine comme étant le personnage auquel réfère le pronom personnel *elle.* Cette identification peut s'effectuer après un calcul interprétatif certes coûteux mais pas impossible, si l'on garde en mémoire qu'il s'agit d'apprentis scripteurs. L'attribution du référent *voisine* à l'expression anaphorique *elle* fait intervenir non seulement « le savoir construit linguistiquement par le texte lui-même », mais aussi « les contenus inférentiels qu'il est possible de calculer à partir des contenus linguistiques » (Reichler-Béguelin, 1988a : 18). Ce calcul inférentiel s'appuie donc sur les éléments déjà saillants du texte, de la situation immédiate et du savoir d'arrière-plan présumé partagé, éléments qui permettent d'accéder au référent visé (cf. chapitre 2, section 2.1.1.2.).

7.2.6. Discordance sémantique

Lorsqu'il s'agit d'attribuer un référent à une expression anaphorique, le sémantisme de l'expression peut être un obstacle à l'interprétation. Dans P2 par exemple, le nom *bruit* possède une valeur dénominative qui lui permet d'évoquer la réalité qui lui est associée (Mortureux, 1998). Mais le référent du nom *bruit* est générique : un bruit peut avoir une multitude de causes et de ce fait, se réaliser, linguistiquement parlant, de diverses manières (onomatopées, SN, phrase complète nécessitant une inférence…). En outre, dans P2, le SN *bruit* est expansé par le syntagme adjectival *grand* (*grand bruit).* L'expansion choisie confère au nom *bruit* une propriété particulière qui doit être prise en compte : un *grand* bruit ne peut être assimilé à n'importe quelle sorte de bruit. Les scripteurs doivent donc gérer à la fois l'introduction du SN démonstratif par un terme linguistique approprié et la cohésion sémantique du terme choisi en accord avec le SN *grand bruit*. Parfois, cette double contrainte n'est réalisée qu'en partie seulement, soulevant ainsi un problème de discordance sémantique.

C'est le cas par exemple dans le texte (31), déjà étudié *supra* pour un autre phénomène.



Camille avait une sœur qui s'appelle Emma. *Elle habitait dans cette maison depuis longtemps.* Il arrivait dans la forêt. Allan se retourna et vit un chat qui miaulait. *Il se retourna en entendant ce grand bruit.* Depuis qu'Allan a vu le chat qui miaulait, il ne va plus dehors la nuit. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 31 : Texte de Romane, CE2 L

Dans ce texte, le référent du SN *bruit* est introduit dans la phrase précédent P2 par le SN expansé *un chat qui miaulait.* Le sémantisme du verbe *miauler* permet de comprendre la nature du bruit entendu par le personnage masculin. Pour cette raison, on peut considérer que le référent du SN démonstratif *ce grand bruit* est introduit. Cependant, comme nous venons de le voir, l'expansion *grand* dans *ce grand bruit* ne permet pas d'accepter n'importe quel type de bruit. Or ici, un chat qui miaule produit certes du bruit mais pas suffisamment pour être qualifié de *grand* bruit.

Au terme de ce chapitre, nous avons exposé nos catégories d'analyse ainsi que les problèmes d'interprétation qui subsistaient. Ce chapitre clôt ainsi la deuxième partie de notre thèse consacrée à la mise en œuvre de l'étude et à la méthodologie de recueil et d'analyse des données. Nous consacrons la partie suivante (partie III) à l'analyse de nos résultats de recherche.

PARTIE III :

Résultats de l'étude

Après avoir posé le cadre théorique de notre recherche (chapitres 1 à 4) et défini la mise en œuvre de notre étude ainsi que la méthodologie d'exploitation de nos données (chapitres 5 à 7), nous allons présenter les résultats[54] obtenus à travers les chapitres 8 à 12. Dans le chapitre 13, nous faisons un état des lieux des principaux résultats de notre recherche.

Chaque chapitre de présentation des résultats est construit selon la même architecture afin de pouvoir être lu séparément. Ce parti pris engendre un effet de répétition pour le lecteur qui choisit de consulter les chapitres les uns après les autres, mais lui donne également un schéma identique sur lequel se reposer étant donné la quantité d'informations délivrées. Ainsi, chaque chapitre présente premièrement une analyse de l'introduction des référents et deuxièmement, une analyse de leur maintien dans le discours.

À la fin de chacun des chapitres, nous proposons un bilan des résultats que nous venons de montrer étape par étape. L'interprétation de ces résultats sera effectuée dans le chapitre de synthèse finale (chapitre 13).

En outre, il nous faut procéder ici à un point terminologique. Dans les sections des chapitres qui analysent le maintien de la référence, nous distinguons d'une part, les reprises des référents à l'aide d'une anaphore (Reichler-Béguelin, 1989 ; cf. chapitre 2), c'est-à-dire deux mentions du référent dans l'ensemble du texte. D'autre part, les reprises des référents dans une chaîne de référence (CDR) (Schnedecker, 1997 ; cf. chapitre 2), c'est-à-dire trois mentions ou plus dans le texte. Nous employons également les termes de *mentions* et *maillons* indifféremment.

Enfin, les illustrations qui sont fournies viennent à l'appui d'un point particulier. Il va de soi que d'autres aspects mériteraient parfois d'être commentés, mais nous avons choisi de ne pas faire de commentaires sur les points fussent-ils remarquables qui ne concernent pas le point particulier que l'on voulait illustrer.

[54] Les tests statistiques *Chi deux* qui auraient permis de mettre en évidence la significativité des différences observées entre les trois niveaux scolaires n'ont pas pu être appliqués dans la mesure où des effectifs ≤ 5 sont présents dans chacun des cas de figure étudiés.

Chapitre 8 : Introduction et maintien du référent du pronom personnel anaphorique *elle*

Dans la tâche-problème que doivent réaliser les élèves de notre étude (cf. partie II, chapitre 6), la première phrase à insérer (P1) est la suivante :

Elle habitait dans cette maison depuis longtemps.

Dans cette phrase, le pronom personnel *elle* nécessite au préalable l'introduction d'un référent car l'emploi d'un pronom implique un degré d'accessibilité fort de l'entité à laquelle il réfère (cf. chapitre 1, partie 1.2.2.). En effet, nous l'avons vu dans le chapitre 1, les pronoms personnels ne comportent pas de tête lexicale. Ils ne portent donc pas en eux-mêmes d'indication sur la catégorie de leur référent, mais cela ne les empêche pas de renvoyer à des particuliers et donc de référer de manière définie (Charolles, 2002). Bien que le pronom personnel ne comporte pas de tête lexicale, comme il est marqué en genre, son emploi engendre la présence d'une source. Dans P1, la seule contrainte exercée par le pronom personnel *elle* est la contrainte de genre : le pronom personnel doit référer à une entité de genre féminin mais cette entité peut ne pas être humaine, ni animée.

Aussi, en 8.1., nous allons étudier la manière dont le référent du pronom personnel *elle* est introduit, en analysant à la fois les formes de la première mention du référent, leurs éventuelles expansions et leurs fonctions syntaxiques. Après avoir été introduit, le référent peut être maintenu tout au long du discours, ou tout du moins dans une partie du discours. En 8.2., nous étudions de ce fait la manière dont le référent est repris tout au long du texte.

8.1. Introduction du référent du pronom personnel *elle*

De manière générale, dans notre corpus, nous observons trois localisations possibles de l'expression référentielle qui introduit le référent (ou *source*), comme le montre le tableau 3 ci-dessous :

- la source est introduite avant la phrase imposée et la référence est maintenue après. Cette stratégie peut être considérée comme experte car lorsqu'un interlocuteur fait

d'une entité son objet de discours, il se limite rarement à la mentionner une seule fois. Il fera donc appel à des expressions qui reprennent l'entité dont il est question (ou *référent*) tout au long du propos et qui sont responsables de la continuité référentielle du texte (Apothéloz, 1995b ; Charolles, 2002 ; cf. chapitre 2, section 2.1).

- la source n'est pas introduite avant la phrase imposée mais elle est localisée après
- la source est absente : la seule mention du référent est celle de la phrase imposée

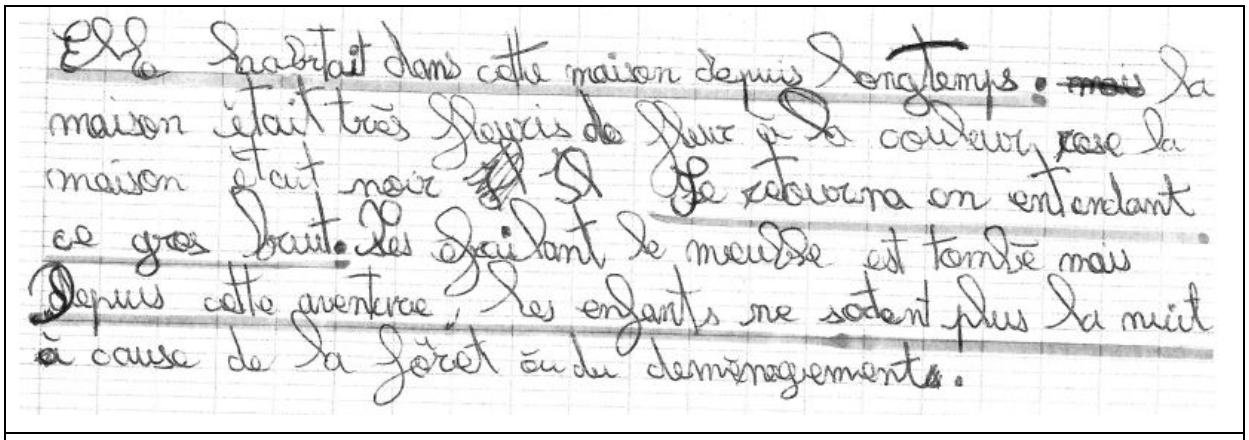
Localisation de la source du pronom personnel *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
avant P1 et maintien après	46	72%	41	73%	59	91%	146	**79%**
après P1	11	17%	11	20%	3	5%	25	**14%**
source absente	7	11%	4	7%	3	5%	**14**	8%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 3 : Localisation de la source du pronom personnel *elle*

Sur l'ensemble du corpus, dans 79% des textes, la source est localisée avant P1 et la référence est maintenue par la suite dans le discours. La fréquence de cette stratégie experte augmente avec le niveau scolaire : 72% au CE2, 73% au CM1, 91% au CM2.

De manière plus marginale, la source est localisée après P1 dans 14% des textes de notre corpus. Il s'agit de cas de *cataphores* (cf. chapitre 2, section 2.1) : le pronom personnel *elle* de P1 est la première mention du référent, qui est repris par la suite sous diverses formes. Cette manière d'introduire le référent décroît globalement avec l'âge des élèves : 17% au CE2, 20% au CM1 et 5% au CM2.

Enfin, sur l'ensemble du corpus, dans 8% des textes (14 textes) la source est absente. C'est-à-dire que la seule mention du référent du pronom personnel *elle* est celle contenue dans P1, comme dans (32). Plus les élèves sont jeunes, plus ils sont nombreux à ignorer le pronom contenu dans P1 : 11% au CE2, 7% au CM1 et 5% au CM2.



Elle habitait dans cette maison depuis longtemps. La maison était très fleurie de fleurs à la couleur rose, la maison était noire. *Il se retourna en entendant ce gros bruit.* Ses (?) le meuble est tombé mais *Depuis cette aventure, les enfants ne sortent plus la nuit.* à cause de la forêt ou du déménagement.

Figure 32 : Texte d'Inès, CE2 U

En d'autres termes, plus le niveau scolaire augmente, plus les textes dans lesquels le référent du pronom personnel *elle* est introduit avant P1 sont nombreux. Cela paraît conforme au développement cognitif de l'élève : il maîtrise de mieux en mieux l'introduction d'un référent dans un texte.

Après avoir dressé une vue d'ensemble des différentes manières de gérer la référence, nous allons étudier les formes de la première mention du référent du pronom personnel *elle*.

8.1.1. Étude des formes de la première mention du référent du pronom personnel *elle*

Nous venons de le voir, le pronom personnel *elle* nécessite une source que les élèves doivent introduire préalablement à P1. Cette source constitue la première mention (M1) du référent du pronom personnel *elle*. Le tableau 4 reproduit les différents types de formes de la première mention du référent féminin *elle.*

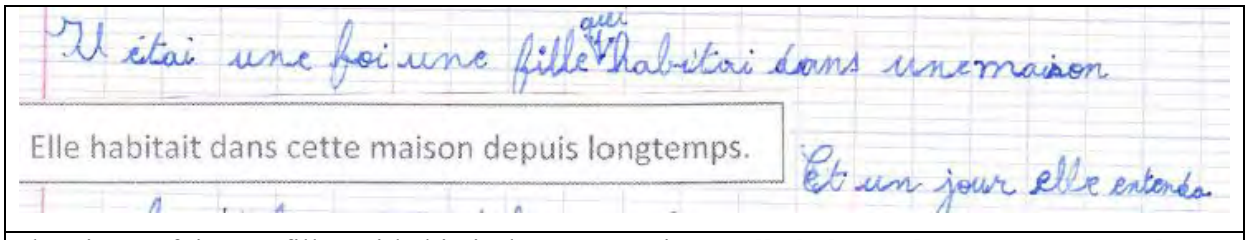
Formes de M1 pour *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN indéfini	31	48%	25	45%	41	63%	97	**52%**
nom propre	10	16%	9	16%	10	15%	29	**16%**
SN défini	0	0%	3	5%	4	6%	7	4%
SN numéral	0	0%	1	2%	0	0%	1	1%
SN possessif	8	**13%**	2	4%	4	6%	14	**8%**
SN démonstratif	1	2%	1	2%	0	0%	2	1%
SN pronominal	7	11%	10	18%	2	3%	19	10%
autre	0	0%	1	2%	1	2%	2	1%
P1 = seule mention	7	11%	4	7%	3	5%	14	8%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 4 : Formes de la première mention du référent du pronom personnel *elle*

Les différentes formes d'introduction du référent de elle

Les travaux qui se sont intéressés à l'introduction des référents dans un discours chez les enfants, notamment ceux de Charolles (1988c ; cf. chapitre 3, section 3.1.2.3.), aboutissent à la conclusion que l'introduction d'un référent par un SN indéfini est la stratégie d'introduction la plus fréquente dans les textes, suivie par le recours à un nom propre. Ces stratégies dominantes sont également celles employées dans les textes narratifs par les scripteurs experts, comme l'ont montré notamment les travaux de Givón (1983 ; 1989) ou ceux de Landragin & *al.* (2015) (cf. chapitre 2, section 2.2.4.). Les autres stratégies restent minoritaires.

Nous retrouvons ces résultats dans notre étude. En effet, l'étude de la forme de la première mention du référent du pronom personnel *elle* montre que, quel que soit le niveau de classe, la première mention est principalement introduite sous la forme d'un SN indéfini : 52% sur l'ensemble du corpus, tel que dans (33) où *elle* de P1 réfère au SN indéfini *une fille* :

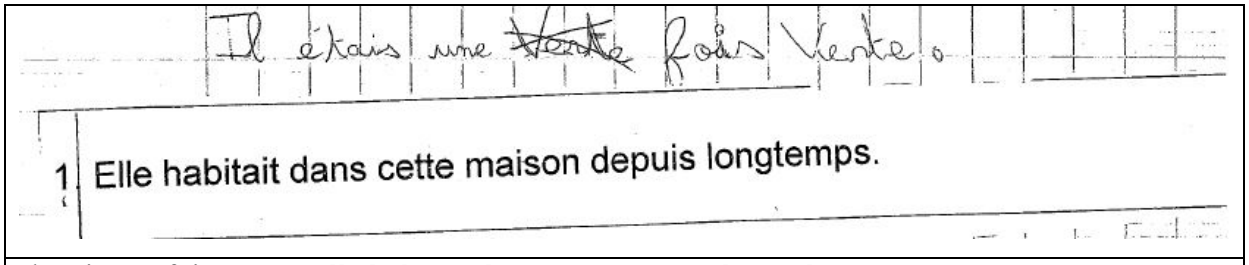


Il était une fois une fille qui habitait dans une maison. *Elle habitait dans cette maison depuis longtemps.*

Figure 33 : Texte de Karl, CE2 T

Le SN indéfini est la forme la plus appropriée pour l'introduction d'un référent. En effet, les expressions nominales indéfinies sont autonomes référentiellement parlant car elles ne demandent pas à ce que l'interlocuteur sache déjà de quoi il va être question dans le propos (cf. chapitre 1, section 1.4.2.1). Les SN indéfinis sont donc utilisés lorsque le locuteur introduit pour la première fois l'objet du discours (cf. chapitre 2, section 2.1.1.3.), ce qui est le cas ici, ou un nouveau référent.

Le deuxième type de forme le plus fréquemment utilisé pour l'introduction du référent du pronom personnel *elle* est le nom propre, comme l'illustre (34) : 16% pour l'ensemble du corpus, dans des proportions quasiment identiques d'un niveau à l'autre.



Il était une fois Verte. *Elle habitait dans cette maison depuis longtemps.*

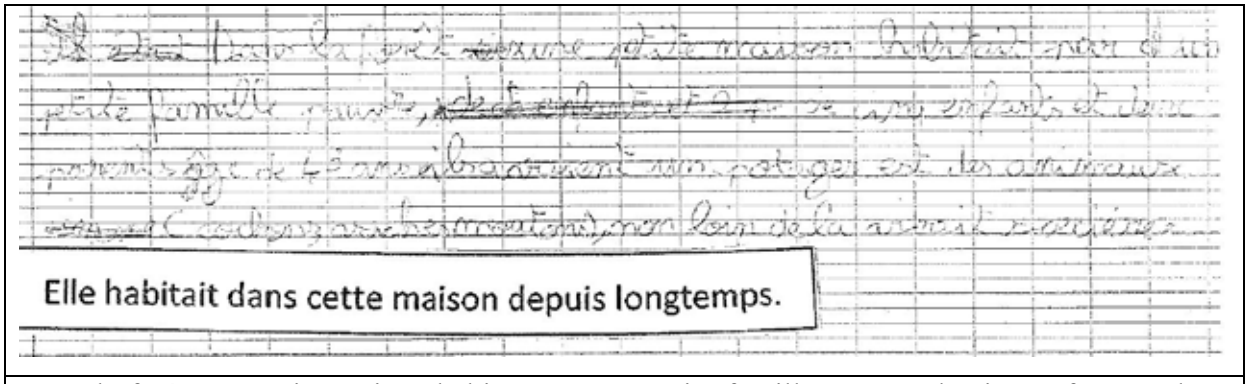
Figure 34 : Texte de Fethallah, CM1 B

De même que le SN indéfini, le nom propre permet d'introduire l'objet du discours. En effet, comme il réfère à un particulier en le discriminant au sein d'une catégorie d'autres particuliers, le nom propre réfère « à des particuliers qui ne sont pas supposés déjà présents dans l'esprit des interlocuteurs » (Charolles, 2002 :74), c'est-à-dire lorsque l'accessibilité du référent est faible (Ariel, 1990 ; cf. chapitre 1, section 1.2.2.). Par cette opération, le nom propre porte l'attention sur une entité non encore saillante et signale qu'elle a de fortes chances de le devenir par la suite.

De manière plus marginale, pour introduire le référent du pronom personnel *elle,* d'autres types de formes linguistiques sont employées. On relève des SN possessifs, tels que

sa femme par exemple : 8% sur l'ensemble du corpus, avec un écart entre le CE2 (13%) et les autres niveaux de classe (4% au CM1, 6% au CM2). On relève également des SN définis (4% des textes) ainsi que des SN démonstratifs ou numéraux (entre 1 et 2 textes pour chaque type de forme).

Dans l'ensemble du corpus, seulement deux textes présentent un type de forme que nous avons qualifié d' « autre » : il s'agit de textes dans lesquels il n'y a pas de déterminant devant le nom qui réfère au pronom personnel *elle,* comme dans l'exemple suivant (35) :



Dans la forêt une petite maison habitée par une petite famille pauvre, de cinq enfants et deux parents âgés de 43 ans. Ils avaient un potager et des animaux (cochons, vaches, moutons). Non loin de là, vivait sorcière. *Elle habitait dans cette maison depuis longtemps.*

Figure 35 : Texte de Thomas, CM1 T1

Étude des formes lexicales de M1 pour elle

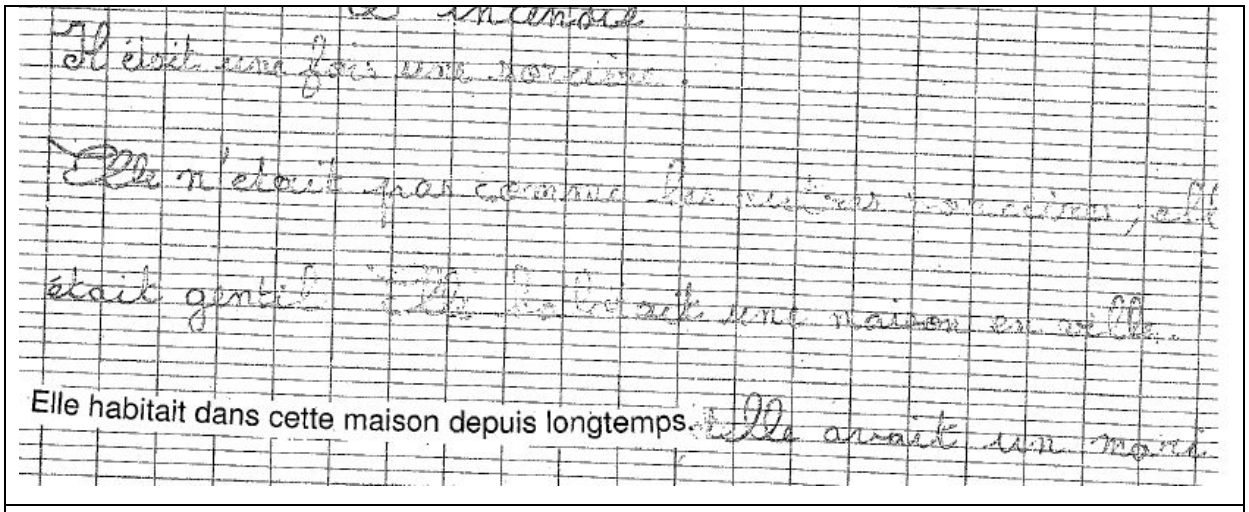
La diversité des formes lexicales employées pour référer au pronom personnel *elle* montre différentes manières de « concevoir » le personnage, comme le montre le tableau 5. Pour en rendre compte, nous nous inspirons de catégories empruntées à Charolles (1988c ; cf. chapitre 3, section 3.1.2.3), auxquelles nous ajoutons les noms relationnels. Ainsi, la référence au personnage féminin peut se réaliser de diverses manières : soit par des dénominations « passe-partout » faisant appel à l'identité du personnage féminin (*fille, dame, femme…)*, soit par des noms relationnels au sein du noyau familial (*mère, grand-mère, cousine…*) soit par des dénominations faisant appel au rôle du personnage dans l'histoire que nous appelons *noms fonctionnels.* Les rôles imposés aux personnages peuvent être hérités de la tradition des contes (*sorcière, bergère, fée, princesse…*) ou non *(maîtresse, voisine…*). Malgré l'absence de contrainte exercée par le pronom personnel *elle* de P1 sur la nature de l'occurrence, aucun texte ne met en scène un référent non humain.

Formes lexicales de M1 pour *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
dénomination « passe-partout »	25	63%	21	64%	34	71%	80	**66%**
nom relationnel	8	20%	4	12%	4	8%	16	13%
nom fonctionnel	7	17%	8	24%	10	21%	25	**21%**
Total	40	100%	33	100%	48	100%	121	100%

Tableau 5 : Formes lexicales de la première mention du référent du pronom personnel elle

L'étude des formes lexicales employées dans les SN (indéfini, défini, possessif, démonstratif, numéral) pour le référent du pronom personnel *elle* montre une prédominance dans l'emploi de dénominations « passe-partout » : 66% des textes. On constate que plus le niveau scolaire augmente, plus la fréquence d'emploi de dénominations *passe-partout* augmente. La forme la plus fréquente est l'unité lexicale *fille* mais on rencontre également *dame, femme, demoiselle, fillette* et *personne*. Le caractère commun, ordinaire des formes lexicales « passe-partout », ne pose pas de problème étant donné qu'elles sont présentes afin de mentionner pour la première fois une entité que l'on ne connaît pas et sur laquelle on n'a aucune information.

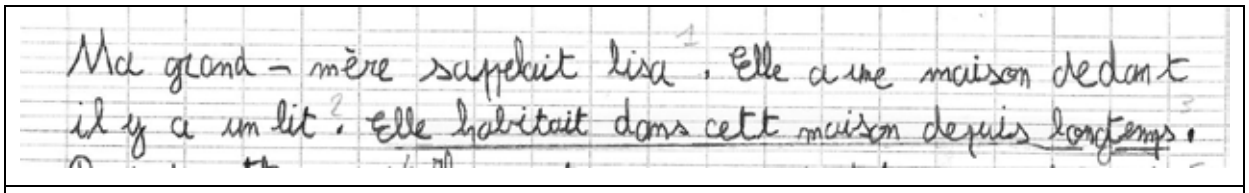
Le deuxième type de forme lexicale le plus fréquent est ce que nous avons nommé *nom fonctionnel* et qui correspond à des formes faisant appel au rôle du personnage dans l'histoire. Ce type de forme concerne 21% des textes, avec un nombre à peu près égal de textes dans chaque niveau. La forme la plus fréquente est l'unité lexicale *sorcière* comme on le voit dans (36), montrant ainsi l'influence de la littérature de jeunesse, en l'occurrence le conte, sur les productions écrites des élèves puisque le personnage de sorcière est un marqueur de genre littéraire (Tauveron, 1995). La fréquence relative de l'emploi de la forme *sorcière* peut être mise en rapport avec la phrase finale P3 qui induit le récit d'une histoire qui doit faire peur. Dans le même registre, on trouve également, dans une moindre mesure, des unités lexicales telles que *princesse, bergère, fée, orpheline*… Dans les formes plus « communes » et non caractéristiques d'un genre littéraire, la forme *voisine* est la plus fréquente, bien que l'on ait aussi *copine, poupée, propriétaire…*



Il était une fois une sorcière. Elle n'était pas comme les autres sorcières, elle était gentille. Elle habitait une maison en ville. *Elle habitait dans cette maison depuis longtemps.*

Figure 36 : Texte de Max, CM2 S

Enfin, les noms relationnels impliquant des liens entre plusieurs personnages au sein du noyau familial sont présents dans 13%. Il s'agit principalement de formes contenues dans des SN possessifs car le déterminant possessif implique un lien entre son référent et un autre personnage. L'unité lexicale la plus employée dans ce cas est *grand-mère*, comme illustré par (37).



Ma grand-mère s'appelait Lisa. Elle a une maison, dedans il y a un lit. *Elle habitait dans cette maison depuis longtemps.*

Figure 37 : Texte d'Adam, CE2 U

Les noms relationnels sont majoritairement employés par les élèves les plus jeunes. Cela peut s'expliquer, comme l'indique Charolles (1988c ; cf. chapitre 3, section 3.1.2.3.), par le fait que plus les élèves sont jeunes, moins ils sont capables de décentration, c'est-à-dire de prendre de la distance vis-à-vis de l'histoire qu'ils racontent. En témoigne la désignation du personnage par son rôle dans la cellule familiale (*grand-mère*, *mère, cousine…),* cellule dans laquelle l'élève évolue au quotidien.

8.1.2. Étude des expansions dans la première mention du référent du pronom personnel *elle*

Lors de l'introduction du référent du pronom personnel *elle* présent dans P1, on trouve dans les textes des SN nus et des SN expansés, présentant un ou plusieurs modifieurs qui sont autant d'ajouts d'informations à propos de l'entité ainsi dénotée (Charolles, 2002).

L'étude des formes de la première mention du référent du pronom personnel *elle* montre globalement que la première mention du référent est expansée dans la moitié des textes, comme l'indique le tableau 6.

Nombre d'expansions de M1 pour *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
1 expansion	21	33%	14	25%	19	29%	54	**29%**
2 expansions	11	17%	10	18%	18	28%	39	**21%**
3 expansions	0	0%	2	4%	1	1%	**3**	2%
pas d'expansion	32	50%	30	53%	27	42%	89	48%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 6 : Nombre d'expansions dans les formes de la première mention du référent du pronom personnel *elle*

À la lecture du tableau, on remarque également que le nombre d'expansions est décroissant : 29% des textes comportent une expansion, 21% en comportent deux et seulement 2 textes comportent trois expansions. En outre, dans notre corpus, nous n'avons pas de textes dans lesquels le SN utilisé pour référer à *elle* comporte plus de 3 expansions. Les résultats montrent aussi que plus les élèves sont jeunes, plus le nombre d'expansions qu'ils emploient est faible.

Étude des types de première expansion dans la première mention du pronom personnel elle

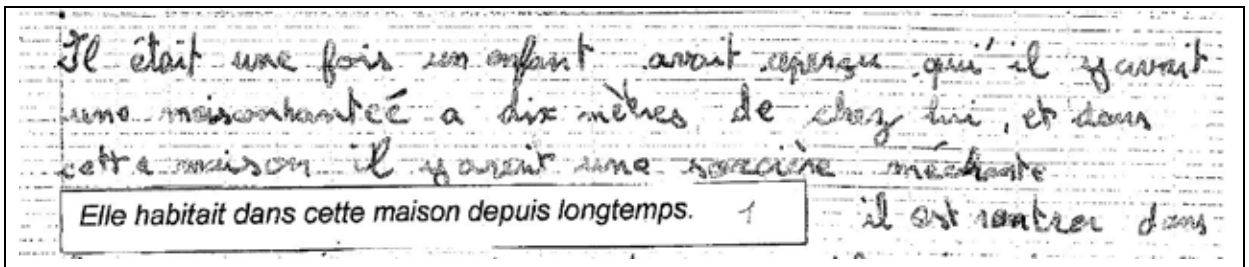
Les types d'expansion utilisés peuvent être de différentes sortes : syntagme adjectival, proposition relative, syntagme prépositionnel ou nom propre.

L'étude des types d'expansions de la première mention du référent du pronom personnel *elle* montre que quel que soit le niveau de classe, lorsque le SN introduisant la première mention est expansé, il l'est majoritairement à l'aide d'un syntagme adjectival : 64% des textes comportant des expansions, comme l'indique le tableau 7.

Type de première expansion de M1 pour *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
syntagme adjectival	15	47%	17	65%	29	76%	61	**64%**
proposition relative	5	16%	6	23%	5	13%	16	**17%**
syntagme prépositionnel	4	13%	2	8%	2	5%	8	**8%**
SN nom propre	8	25%	1	4%	2	5%	11	**11%**
Total	32	100%	26	100%	38	100%	96	100%

Tableau 7 : Nombre d'expansions dans les formes de la première mention du référent du pronom personnel *elle*

Par exemple dans (38), le SN indéfini *une sorcière* est expansé à l'aide du syntagme adjectival *méchante* :

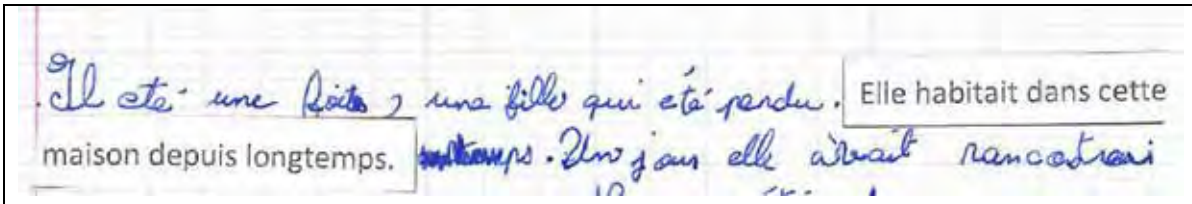


Il était une fois, un enfant avait aperçu qu'il y avait une maison hantée à dix mètres de chez lui et dans cette maison, il y avait une sorcière méchante. *Elle habitait dans cette maison depuis longtemps.*

Figure 38 : Texte de Yassine, CM2B

Dans une moindre mesure, les expansions peuvent également être :

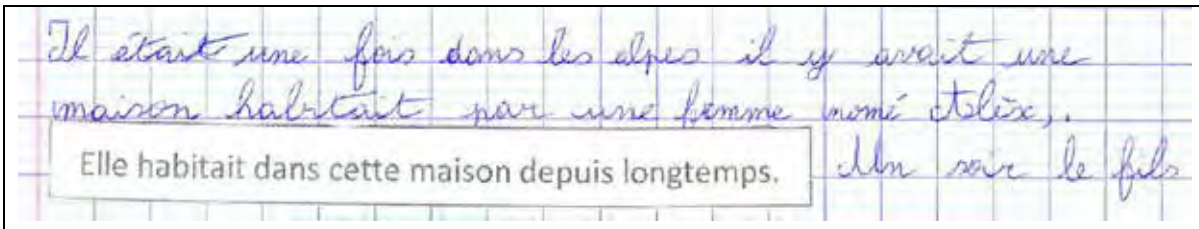
- une **proposition relative** : 17% des textes, comme dans (39) où la proposition relative *qui était perdue* est un modifieur du SN indéfini *une fille* :



Il était une fois une fille qui était perdue. *Elle habitait dans cette maison depuis longtemps.*

Figure 39 : Texte de Loulie, CE2 P

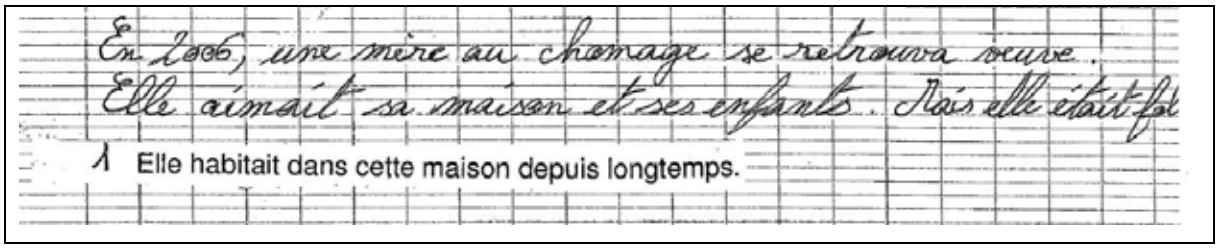
- **un nom propre** : 11% des textes avec un net écart entre les niveaux : 25% au CE2 contre 4% et 5% au CM1 et CM2. Dans (40) par exemple, le SN indéfini *une femme* est suivi du nom propre *Alix* :



Il était une fois dans les Alpes, il y avait une maison habitée par une femme nommée Alix. *Elle habitait dans cette maison depuis longtemps.*

Figure 40 : Texte de Victor, CE2 T

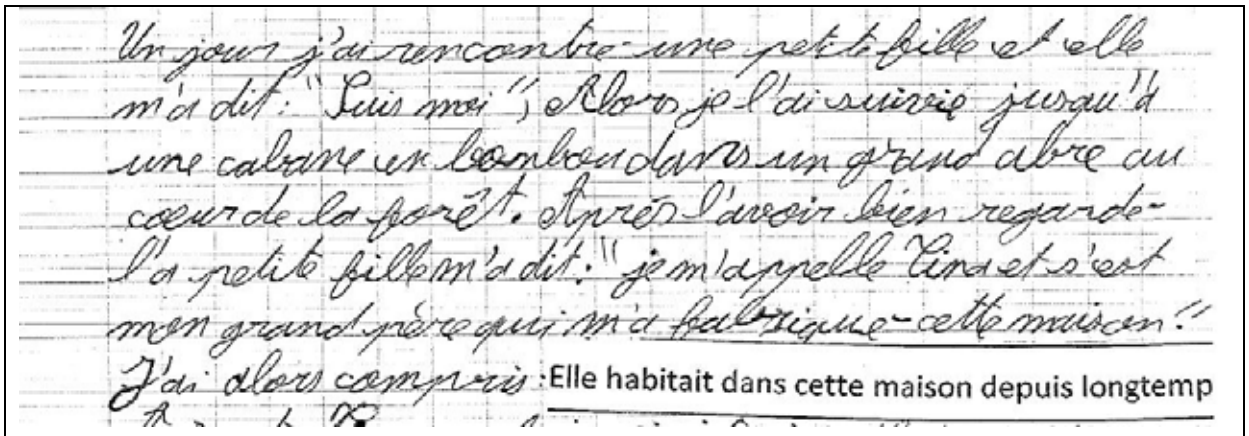
- un **SN prépositionnel** : 8% des textes, avec là encore, un écart entre les différents niveaux scolaires. Il s'agit d'expansions comme *au chômage* dans (41) :



En 2006, une mère au chômage se retrouva veuve. Elle aimait sa maison et ses enfants. Mais elle était folle. *Elle habitait dans cette maison depuis longtemps.*

Figure 41 : Texte de Maxime, CM2 S

Toutefois, il nous faut nuancer les résultats obtenus en ce qui concerne le nombre de textes dans lesquels le SN est expansé par un SN adjectival (cf. tableau 7). En effet, parmi tous les SN indéfinis comportant le nom *fille*, 52% comportent le SN adjectival *petite* (*petite fille*) et 20 % le SN adjectival *jeune* (*jeune fille*). On constate donc un grand nombre d'associations privilégiées de deux termes, c'est-à-dire l'assemblage d'un nom avec un adjectif qualificatif qui lui est souvent associé, tel que *petite fille* (cf. 42)*, jeune fille, vieille dame,* fréquents dans la littérature de jeunesse. Il s'agit alors de collocation (Bally, 1909 ; Tutin & Grossman, 2002).


Un jour j'ai rencontré une petite fille et elle m'a dit : "Suis moi", Alors je l'ai suivie jusqu'à une cabane en bambou dans un grand arbre au coeur de la forêt. Après l'avoir bien regardé-l'a petite fille m'a dit : "je m'appelle Tina et c'est mon grand père qui m'a fabrique cette maison !" J'ai alors compris : Elle habitait dans cette maison depuis longtemp

Un jour, j'ai rencontré une petite fille et elle m'a dit : « Suis-moi ». Alors je l'ai suivie jusqu'à une cabane en bambou dans un grand arbre au cœur de la forêt. Après l'avoir bien regardé, la petite fille m'a dit : « je m'appelle Tina et c'est mon grand-père qui m'a fabriqué cette maison ». J'ai alors compris. *Elle habitait dans cette maison depuis longtemps.*

Figure 42 : Texte de Naélia, CM1 T1

Étude des types de deuxième expansion dans la première mention du pronom personnel elle

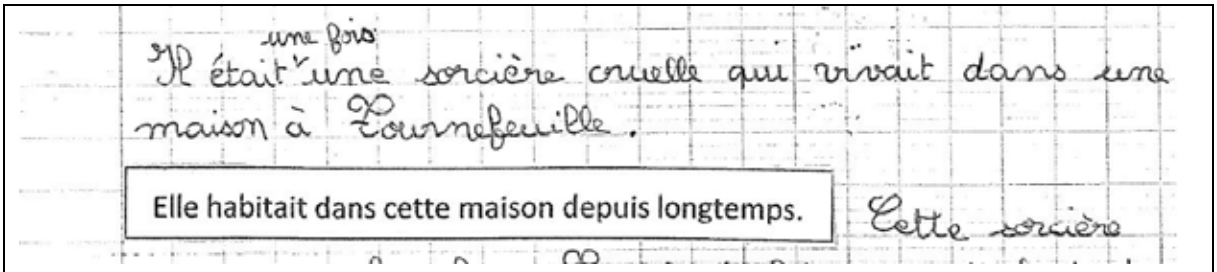
L'étude du type de deuxième expansion de la première mention du référent du pronom personnel *elle* montre que quel que soit le niveau de classe, la proposition relative est le type de deuxième expansion le plus fréquent : 71 % des textes.

Type de deuxième expansion de M1 pour *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
proposition relative	8	73%	7	58%	15	79%	30	71%
syntagme prépositionnel	2	18%	2	17%	2	11%	6	14%
SN nom propre	1	9%	3	25%	2	11%	6	14%
Total	11	100%	12	100%	19	100%	42	100%

Tableau 8 : Type de deuxième expansion dans les formes de la première mention du référent du pronom personnel *elle*

Lorsqu'on regarde de plus près la composition des propositions relatives employées, on remarque qu'elles peuvent être de trois sortes :

- des propositions relatives contenant un prédicat dont *maison* est un argument. Le SN démonstratif *cette maison*, qui est présent dans P1, correspond à la deuxième occurrence à introduire. Le prédicat peut être *habiter* ou *vivre* comme dans (43) ci-dessous, *qui vivait dans une maison à Tournefeuille* :

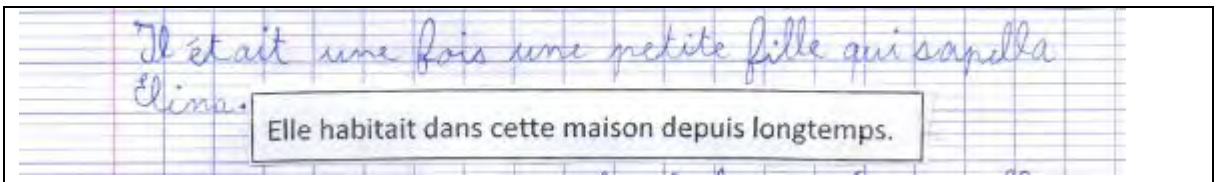


Il était une fois une sorcière cruelle qui vivait dans une maison à Tournefeuille. *Elle habitait dans cette maison depuis longtemps.*

Figure 43 : Texte d'Emma, CM1 T1

Ce type de proposition relative démontre un certain degré de maîtrise de l'introduction de référents. En effet, par ce moyen, les élèves anticipent déjà, dans une certaine mesure, l'introduction du référent de l'occurrence *cette maison* présente dans P1.

- des propositions relatives de dénomination, c'est-à-dire comportant un nom propre, caractérisées par la présence des verbes *nommer* ou *s'appeler*, comme l'illustre (44) suivant avec la proposition relative *qui s'appelait Elina* :

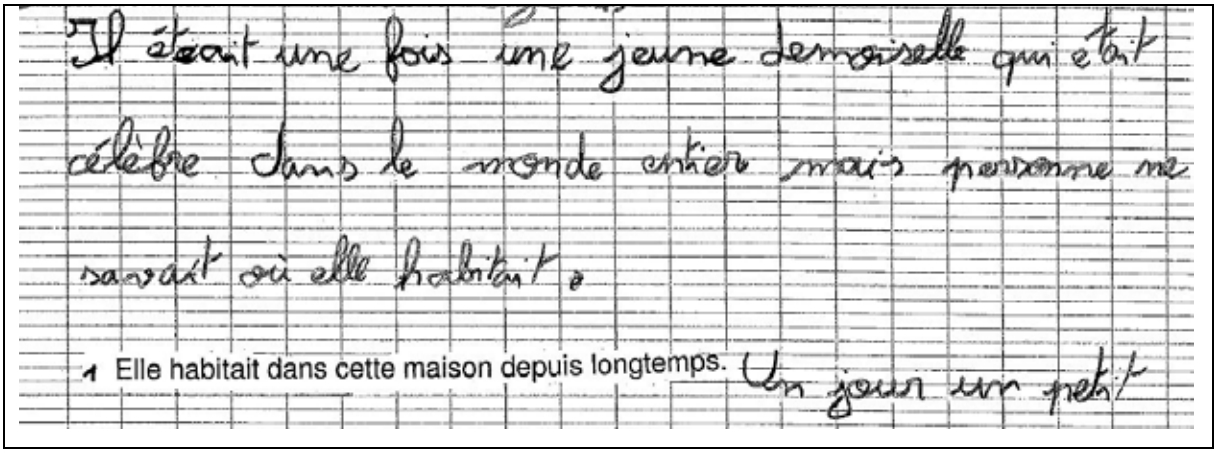


Il était une fois une petite fille qui s'appelait Elina. *Elle habitait dans cette maison depuis longtemps.*

Figure 44 : Texte de Julian, CE2 L

L'utilisation d'une proposition relative comportant un nom propre est un atout dans la construction de la référence. En effet, ainsi introduit, le nom de baptême du personnage pourra être réutilisé dans la suite du texte, lorsqu'il sera nécessaire de réinstancier le personnage.

- des propositions relatives apportant d'autres types d'informations sur le personnage, comme par exemple le fait qu'il est célèbre dans (45) :



Il était une fois une jeune demoiselle qui était célèbre dans le monde entier mais personne ne savait où elle habitait. *Elle habitait dans cette maison depuis longtemps.*

Figure 45 : Texte de Antonin, CM2 S

Ce type d'information introduit à l'aide d'une proposition relative n'est pas utile à la référence au personnage, mais il apporte des précisions sur celui-ci et permet d'enrichir le texte.

Le tableau 9 ci-dessous rapporte la fréquence de chaque type de propositions relatives que nous venons de détailler.

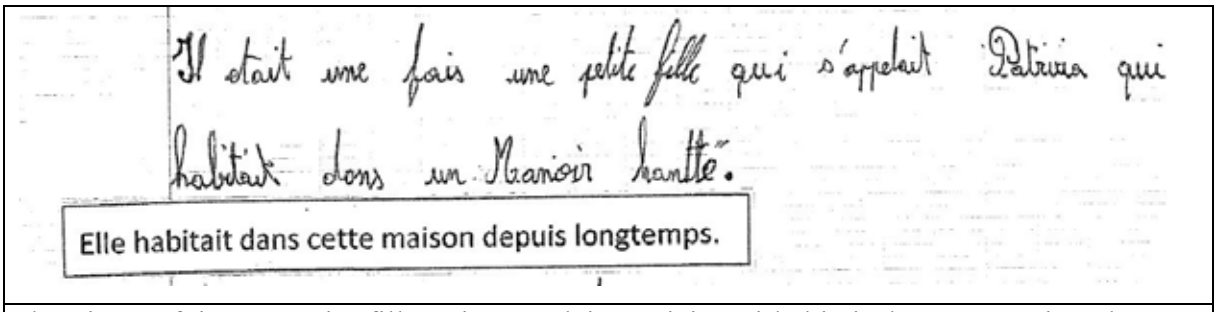
Composition des propositions relatives en expansion 2 pour M1 de *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
PR contenant un prédicat dont *maison* est un argument	2	25%	2	29%	6	40%	10	33%
PR dénominative	6	75%	4	57%	5	33%	15	50%
Autre caractéristique du personnage	0	0%	1	14%	4	27%	5	17%
Total	8	100%	7	100%	15	100%	30	100%

Tableau 9 : Composition des propositions relatives présentes dans la deuxième expansion de la première mention du référent du pronom personnel *elle*

Étude des types de troisième expansion dans la première mention du pronom personnel elle

Parmi les textes comportant des expansions dans la première mention du référent du pronom personnel *elle*, il y en a seulement deux qui ont ajouté trois expansions.

Dans (46), le SN référant à *elle* est expansé premièrement par l'adjectif qualificatif *petite*, puis par une première proposition relative incluant un nom propre *qui s'appelait Patricia,* suivie d'une seconde proposition relative dont le verbe est en rapport de sens avec l'occurrence *cette maison*, *qui habitait dans un manoir hanté* :


Il était une fois une petite fille qui s'appelait Patricia qui habitait dans un Manoir hanté.

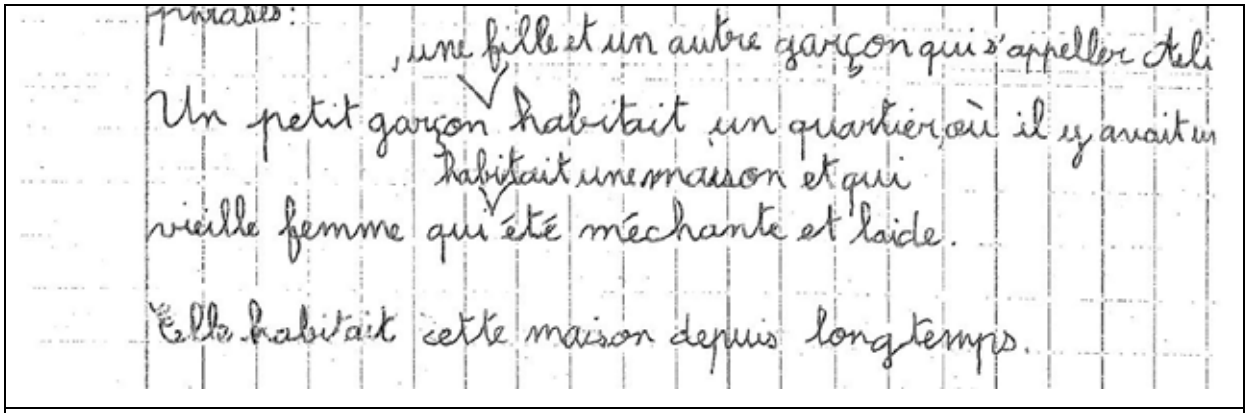
Elle habitait dans cette maison depuis longtemps.

Il était une fois une petite fille qui s'appelait Patricia qui habitait dans une maison hantée. *Elle habitait dans cette maison depuis longtemps.*

Figure 46 : Texte de Lucas 2, CM1 T1

Par la deuxième expansion du SN, *qui s'appelait Patricia*, l'élève apporte un élément important pour le maintien de la référence à ce personnage dans la suite de son texte. En effet, le fait de baptiser le personnage permettra à l'élève d'utiliser le nom propre pour référer à ce personnage lorsqu'il sera nécessaire de le réinstancier par la suite dans son texte. La troisième expansion quant à elle, *qui habitait dans un manoir hanté*, montre que l'élève prépare déjà l'introduction de l'occurrence *cette maison*, présente en P1.

Dans (47) ci-après, contenant plusieurs ajouts après révision du texte, le pronom personnel *elle* a pour source le SN indéfini *une femme*, expansé à l'aide de l'adjectif qualificatif *vieille*, puis d'une première proposition relative, *qui habitait une maison,* coordonnée à une seconde proposition relative, *qui était méchante et laide* :


Un petit garçon, une fille et un autre garçon qui s'appeller Ali habitait un quartier où il y avait un vieille femme qui habitait une maison et qui été méchante et laide.

Elle habitait cette maison depuis longtemps.

Un petit garçon, une fille et un autre garçon qui s'appelait Ali habitaient un quartier où il y avait une vieille femme qui habitait une maison et qui était méchante et laide. *Elle habitait dans cette maison depuis longtemps.*

Figure 47 : Texte de Benjamin, CM2 T

Dans ce texte, on voit que l'élève avait d'abord expansé le SN *une vieille femme* à l'aide de la proposition relative *qui était méchante et laide*, proposition relative qui apporte des

informations sur le personnage mais n'est pas utile à la référence. Dans un second temps d'écriture, l'élève a ajouté la proposition relative *qui habitait une maison*. Cet ajout montre que la nécessaire introduction de l'occurrence *cette maison*, présente dans P1, a bien été prise en compte par l'élève mais lors de la révision de son texte, à l'inverse de l'exemple commenté ci-dessus où l'élève a géré les deux éléments dans un même temps d'écriture.

Après avoir analysé les expansions contenues dans la première mention du référent du pronom personnel *elle*, nous nous intéressons aux fonctions syntaxiques occupées par les différentes formes de la première mention du référent.

8.1.3. Étude des fonctions syntaxiques des formes de M1

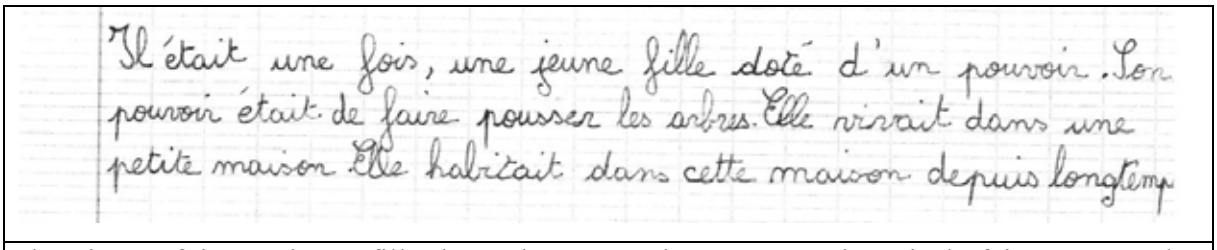
Nous l'avons déjà vu, le pronom personnel *elle* de P1 nécessite l'introduction préalable d'un référent. La première mention de ce référent constitue donc une information totalement nouvelle puisqu'elle n'est pas encore connue du locuteur. Cette information représente également le thème de l'énoncé (cf. chapitre 1, section 1.2.3.), qui peut se réaliser sous la forme de sujet syntaxique et se retrouver en position initiale dans la proposition, suivant la stratégie du sujet thématique (Karmiloff-Smith, 1979, cf. chapitre 3, section 3.1.). En outre, le pronom personnel *elle* de P1 est placé en position de sujet syntaxique. Les textes devraient donc présenter la même « hiérarchie » que dans P1.

En accord avec ces principes, l'étude des fonctions syntaxiques des formes de la première mention du référent du pronom personnel *elle* montre que la fonction sujet est la plus fréquente dans les textes : 39% sur l'ensemble du corpus, bien que sa fréquence d'emploi tend à diminuer lorsque le niveau de classe diminue, comme l'indique le tableau 10. Mais la construction présentative apparaît également très fréquemment.

Fonctions syntaxiques de M1 pour *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
sujet	29	45%	25	45%	19	29%	73	**39%**
construction présentative	17	27%	16	29%	31	48%	64	**35%**
phrase nominale	1	2%	0	0%	2	3%	3	2%
complément du nom	1	2%	0	0%	0	0%	1	1%
complément de verbe	5	8%	6	11%	7	11%	18	10%
complément de phrase	1	2%	2	4%	0	0%	3	2%
autre	3	5%	3	5%	3	5%	9	5%
P1 = seule mention	7	11%	4	7%	3	5%	14	8%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 10 : Types de fonctions syntaxiques occupées par la première mention du référent du pronom personnel *elle*

À la lecture du tableau, on constate en effet que les constructions présentatives sont le deuxième type de structure syntaxique le plus fréquent : 35% des textes du corpus. Il s'agit de constructions comme *Il était une fois*, caractéristiques des contes, ou bien de constructions typiques de l'oral, telles que *c'était / il y a*. Par leur construction, ces différents types d'expressions permettent de mettre une information en position saillante (cf. chapitre 1, section 1.2.4.), comme l'illustre l'exemple (48). Notons tout de même que l'emploi de constructions présentatives augmente en fonction du niveau de classe.


Il était une fois, une jeune fille doté d'un pouvoir. Son pouvoir était de faire pousser les arbres. Elle vivait dans une petite maison. Elle habitait dans cette maison depuis longtemp

Il était une fois une jeune fille dotée d'un pouvoir. Son pouvoir était de faire pousser des arbres. Elle vivait dans une petite maison. *Elle habitait dans cette maison depuis longtemps.*

Figure 48 : Texte de Quentin, CM2 M

8.2. Maintien du référent du pronom personnel *elle*

Après avoir été introduit, le référent du pronom personnel *elle* peut être maintenu plus ou moins longtemps dans le texte. En effet, en tant que personnage saillant, on s'attend à ce que le référent du pronom personnel *elle* soit mentionné plusieurs fois dans le texte à l'aide de diverses expressions référentielles, constituant ainsi une chaîne de référence (CDR) à la *texture* (Halliday & Hasan, 1976 ; cf. chapitre 1, section 1.3.) importante.

Aussi, en 8.2.1., nous dressons une vue générale du maintien de la référence à travers l'étude du nombre de mentions des CDR et de leur longueur moyenne. En 8.2.2., nous étudions de manière plus précise différents cas de figure, selon qu'il s'agit d'anaphore (1 mention + P1 soit 2 mentions au total), de CDR minimales (2 mentions + P1 soit 3 mentions au total) ou de CDR plus longues (3 mentions ou plus + P1 soit 4 mentions ou plus au total).

8.2.1. Vue d'ensemble des CDR incluant le pronom personnel *elle*

De manière générale, l'étude du nombre de mentions (ou maillons) de la CDR incluant le pronom personnel *elle* montre que la CDR la plus longue comporte 20 maillons + celui compris dans P1, soit 21 maillons au total, et se trouve au CM1, comme l'illustre le tableau 11. On constate également que les CDR les plus longues (14, 15 ou 16 mentions + P1) sont représentées dans chacun des niveaux de classe mais par un seul texte à chaque fois.

Nombre de mentions dans la CDR de *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
1 mention + P1	11	17%	13	23%	8	12%	32	**17%**
2 mentions + P1	13	20%	10	18%	11	17%	34	**18%**
3 mentions + P1	10	16%	6	10%	11	17%	27	15%
4 mentions + P1	8	13%	3	5%	5	7%	16	**8%**
5 mentions + P1	5	8%	4	7%	5	7%	14	7%
6 mentions + P1	3	5%	3	5%	8	12%	14	7%
7 mentions + P1	0	0%	5	9%	2	3%	7	4%
8 mentions + P1	0	0%	4	7%	7	10%	11	6%
9 mentions + P1	3	5%	1	2%	1	2%	5	**3%**
10 mentions + P1	0	0%	1	2%	1	2%	2	1%
11 mentions + P1	1	2%	0	0%	1	2%	2	1%
12 mentions + P1	1	2%	1	2%	0	0%	2	1%
14 mentions + P1	0	0%	0	0%	1	2%	1	1%
15 mentions + P1	1	2%	0	0%	0	0%	1	1%
16 mentions + P1	0	0%	0	0%	1	2%	1	1%
20 mentions + P1	0	0%	1	2%	0	0%	1	1%
P1 = seule mention	7	11%	5	9%	3	5%	15	8%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 11 : Nombre de mentions dans la CDR du pronom personnel *elle*

De manière globale, on constate que les textes comportant une CDR minimale (2 mentions + P1, soit 3 mentions au total), sont les plus nombreux : 18% du corpus. La longueur moyenne des CDR incluant le pronom personnel *elle* semble contredire ce résultat, hormis pour le CE2, puisque le nombre moyen de maillons par CDR est de :

- 3,7 maillons au CE2
- 5,4 maillons au CM1 et au CM2

Cependant, le nombre moyen de maillons par CDR est à nuancer car le nombre de CDR longues au CM1 et au CM2 vient influencer la longueur moyenne. On peut tout de même affirmer que le nombre d'expressions coréférentielles employées augmente avec le niveau de classe.

L'emploi d'anaphores (1 mention + P1 soit 2 mentions au total) représente 17% des textes du corpus. Autrement dit, en proportion, il y a presque autant d'élèves qui mentionnent le référent de manière anaphorique (2 mentions) qu'à l'aide d'une CDR minimale (3 mentions).

Les textes présentant des CDR plus longues (4 à 9 mentions) sont moins nombreux : entre 8% et 3% du corpus et leur nombre décroît quand la longueur de la chaîne augmente. Les textes présentant de très longues CDR (10 mentions et au-delà) sont minoritaires mais on en trouve dans tous les niveaux de classe.

8.2.2. Différentes configurations des CDR

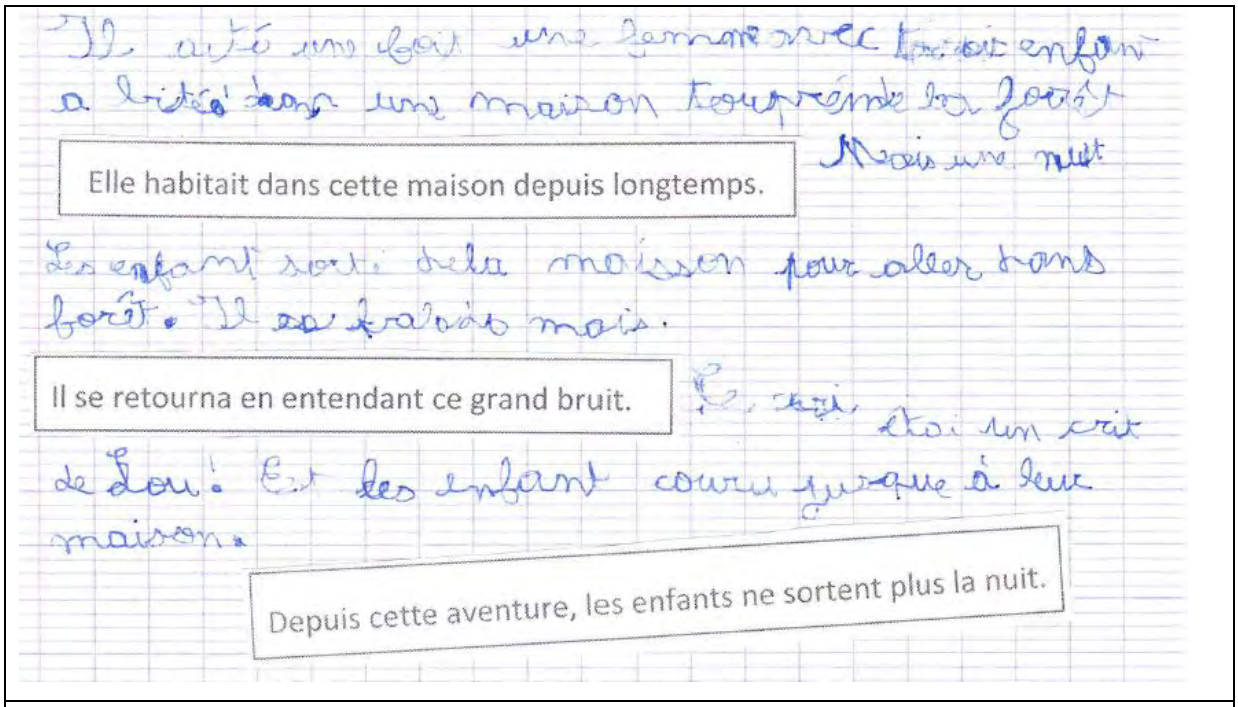
Nous l'avons vu, notre corpus regroupe plusieurs configurations de CDR. Il peut s'agir d'anaphore ou de cataphore (1 mention + P1), de CDR minimales (2 mentions + P1) ou de CDR plus longues (3 mentions ou plus + P1). Chacune de ces configurations fait l'objet d'une étude plus précise afin de savoir si le nombre de mentions d'un référent influe sur les caractéristiques des expressions référentielles utilisées.

8.2.2.1. Étude des cas d'anaphore ou de cataphore : 1 mention + P1

Les textes contenant 1 mention du référent + P1 (soit 2 mentions au total) représentent 17% des textes du corpus (32 textes). Dans cette configuration, on distingue les cas d'anaphore (68%, soit 22 textes) et les cas de cataphore (31%, soit 10 textes).

On parle de cas d'anaphore lorsqu'il y a une mention du référent avant P1 et celle déjà présente dans P1 (Reichler-Béguelin, 1988 ; cf. chapitre 2). Les caractéristiques de la première mention du référent sont alors les mêmes que celles analysées dans la section 8.1. lorsque nous avons étudié l'introduction du référent : la forme linguistique la plus fréquente est le SN indéfini en position de sujet syntaxique ou de construction présentative.

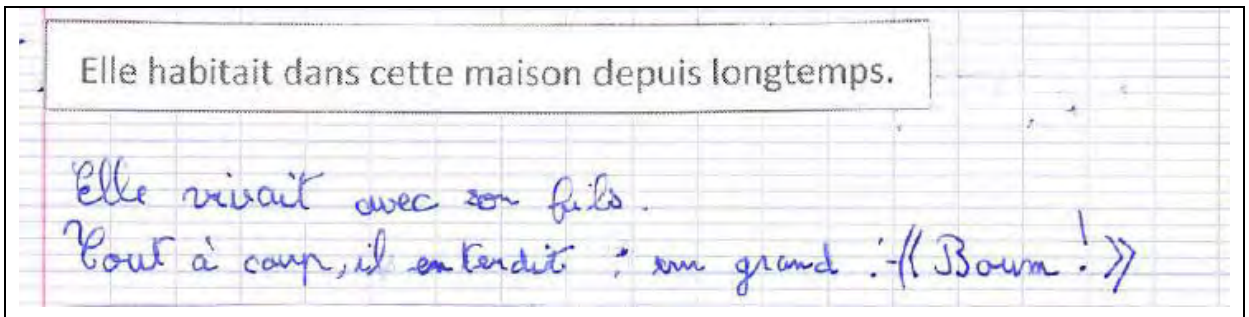
Les cas d'anaphore peuvent parfois s'expliquer par la brièveté du texte : lorsque le texte est très court, il y a moins de mentions des référents. Les cas d'anaphore peuvent également témoigner d'une gestion de la référence au pronom personnel plutôt locale ou « en arrière-plan ». En effet, le petit nombre de mentions du référent ne permet pas d'en faire une entité véritablement présente dans le discours ou bien de construire une histoire entière dont elle serait le sujet thématique. Le personnage féminin n'est finalement qu'un personnage secondaire, comme dans (49) :



Il était une fois une femme avec trois enfants habitait dans une maison tout près de la forêt. *Elle habitait dans cette maison depuis longtemps.* Mais une nuit, les enfants sortirent de la maison pour aller dans la forêt. Il se baladait mais. *Il se retourna en entendant ce grand bruit.* Ce cri était un cri de loup ! Et les enfants coururent jusqu'à leur maison.

Figure 49 : Texte de Nicolas, CE2 T

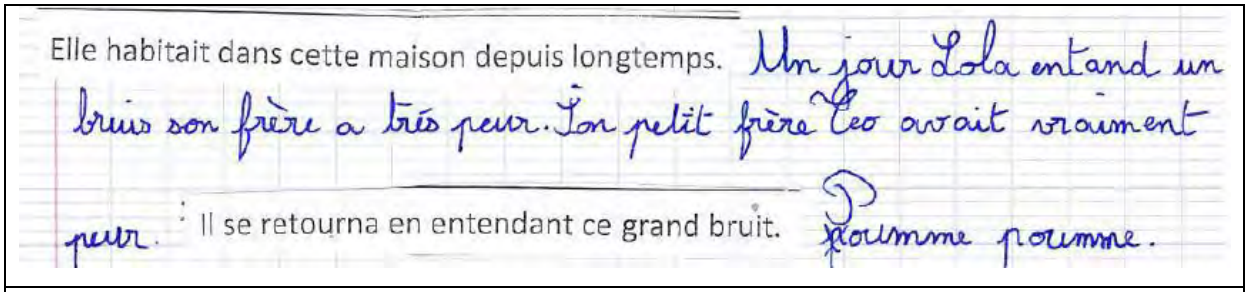
On parle de cataphore (cf. chapitre 2, section 2.1.) lorsque le pronom personnel *elle* de P1 est la première mention du référent. La source du référent est alors localisée après P1, comme dans (50) avec le pronom personnel *elle* de la deuxième phrase.



Elle habitait dans cette maison depuis longtemps. Elle vivait avec son fils. Tout à coup, il entendit un grand « Boum ! ».

Figure 50 : Texte d'Arthur, CM1 T2

La première mention du référent étant situé après P1, on s'attend à ce qu'elle apparaisse principalement sous la forme d'un SN défini, car le référent est devenu plus accessible (cf. chapitre 1, section 1.2.). Pourtant, les résultats de notre étude montrent que le pronom personnel *elle,* exclusivement en position de sujet syntaxique, est la forme employée dans tous les cas de cataphore à l'exception d'un texte, dans lequel la source du référent prend la forme d'un nom propre (51), conférant au texte un autre degré de cohérence. En effet, l'emploi d'un pronom personnel pour la source du référent qui est lui-même localisé dans P1 sous la forme d'un pronom personnel fait qu'on n'arrive pas à « caractériser » le personnage, à lui attribuer des traits qui permettent de le préciser dans notre représentation mentale. L'emploi du nom propre en revanche, même si on ne s'y attend pas, apporte un degré d'information supplémentaire qui permet une construction plus précise.


Elle habitait dans cette maison depuis longtemps. Un jour Lola entand un bruis son frère a très peur. Son petit frère Téo avait vraiment peur. Il se retourna en entendant ce grand bruit. Poumme poumme.

Elle habitait dans cette maison depuis longtemps. Un jour Lola entend un bruit, son frère a très peur. Son petit frère Théo avait vraiment peur. *Il se retourna en entendant ce grand bruit.* « Poum, poum ».

Figure 51 : Texte d'Emmaëlle, CE2 P

En outre, l'étude des fonctions syntaxiques de la première mention du référent du pronom personnel *elle* dans les cas de cataphore montre qu'il apparaît exclusivement en position de sujet syntaxique.

8.2.2.2. Étude des CDR minimales : 2 mentions + P1

Nous l'avons vu, le cas de figure le plus fréquent dans notre corpus lorsqu'il s'agit de maintenir la référence à l'occurrence *elle* est le recours à la configuration suivante : 2 mentions + P1 (cf. tableau 10 *supra*), c'est-à-dire 3 mentions au total. Il s'agit là de la configuration minimale dans laquelle on peut parler de CDR (Schnedecker, 1997 ; cf. chapitre

2, section 2.2.). On parle donc de CDR minimale, en opposition avec les CDR contenant plus de 3 maillons que nous aborderons plus tard.

Les CDR minimales témoignent d'une gestion de la référence plus globale. En effet, en mentionnant le référent du pronom personnel *elle* plusieurs fois en dehors de P1, l'élève commence à construire son personnage en lui octroyant une place plus importante dans son récit. Les travaux qui étudient le maintien de la référence dans des textes narratifs de scripteurs experts, notamment ceux de Landragin & *al.* (2015 ; cf. chapitre 2, section 2.2.4), montrent qu'une CDR prototypique se réalise le plus souvent sous la forme :

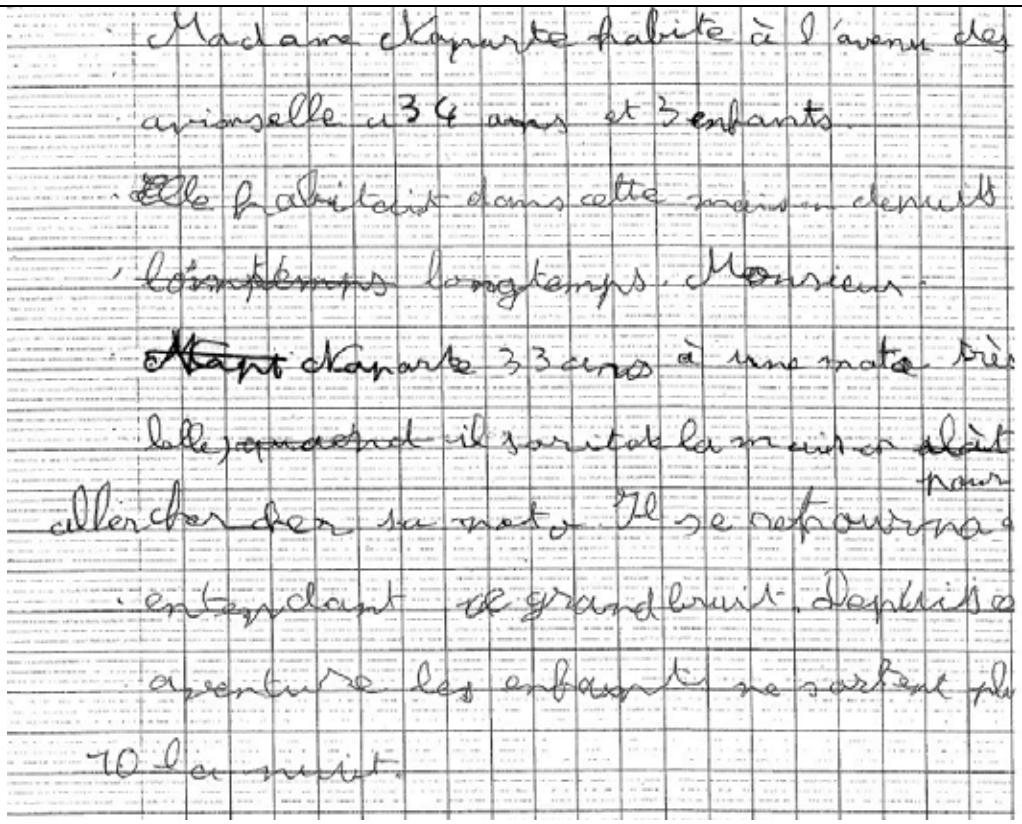
> SN indéfini ou nom propre – description définie – successions de pronoms personnels de troisième personne.

Dans les cas de CDR minimales que nous étudions ici (2 mentions + P1 soit 3 mentions au total), le pronom personnel est présent dans P1. On peut donc s'attendre à ce que la deuxième mention du référent soit majoritairement un SN défini lorsque P1 est placé en troisième et dernière position dans la CDR. Lorsque P1 est placé en deuxième position dans la CDR, on devrait alors trouver un pronom en troisième mention.

L'étude de la deuxième mention du référent dans les CDR minimales montre que la forme la plus fréquente, quelle que soit la position de P1, est le pronom : 44% des textes contenant une CDR minimale, comme le montre le tableau 12 et l'exemple (52) :

Formes de M2 dans les CDR minimales de *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN défini	1	8%	4	**40%**	2	18%	7	20%
SN pronominal	6	46%	2	20%	7	64%	15	**44%**
SN possessif	3	23%	1	10%	0	0%	4	12%
SN démonstratif	0	0%	0	0%	2	18%	2	6%
SN nom propre	3	23%	3	30%	0	0%	6	18%
Total	13	100%	10	100%	11	100%	34	100%

Tableau 12 : Formes de la deuxième mention du référent du pronom personnel *elle* dans les CDR minimales



Madame Naparte habite à l'avenue des avions, elle a 34 ans et 3 enfants. *Elle habitait dans cette maison depuis longtemps.* Monsieur Naparte, 33 ans, a une moto très belle. Il sortit de la maison pour aller chercher sa moto. *Il se retourna en entendant ce grand bruit. Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 52 : Texte de Nicolas, CM2 T

Malgré le petit nombre de textes concernés, on constate tout de même que si la prédominance du pronom se retrouve dans les niveaux du CE2 et du CM2, le SN défini est majoritaire au CM1 : il est présent dans 40% des cas, quelle que soit la position de P1.

L'étude des fonctions syntaxiques lors de l'introduction du référent (cf. ce chapitre, section 8.1.3.) a montré que la première mention du référent est majoritairement en position de sujet syntaxique. Le personnage féminin de P1 étant une entité saillante, on s'attend à ce qu'il le reste tout au long de la CDR : la deuxième mention du référent devrait également se trouver en position de sujet syntaxique.

Cette hypothèse est confirmée par nos résultats : la deuxième mention des CDR minimales est principalement en position de sujet syntaxique (76% des textes concernés), comme le montre le tableau 13 et l'illustre l'exemple précédent (52).

Fonctions syntaxiques de M2 dans les CDR minimales de *elle*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
sujet	8	61%	7	70%	11	100%	26	76%
complément de verbe	4	31%	2	20%	0	0%	6	18%
complément du nom	1	8%	1	10%	0	0%	2	6%
Total	13	100%	10	100%	11	100%	34	100%

Tableau 13 : Fonctions syntaxiques de la deuxième mention du référent du pronom personnel *elle* dans les CDR minimales

8.2.2.3. Étude des CDR plus longues : 3 mentions (ou plus) + P1

Dans les textes comportant des CDR plus longues, c'est-à-dire 3 mentions (ou plus) + celle contenue dans P1, nous rappelons que nous pouvons distinguer (cf. tableau 11) :

- Les CDR comportant 3 mentions + P1 (soit 4 mentions au total), qui représentent 15% des textes du corpus
- Les CDR comportant entre 4 et 9 mentions + P1, qui représentent entre 8 et 3% du corpus
- Les CDR comportant entre 10 et 16 mentions + P1, qui représentent entre 1 et 2 textes chacune et qui figurent dans chaque niveau de classe
- La CDR la plus longue de notre corpus qui comporte 20 mentions + P1 et qui se trouve au CM1

Nous ne pourrions ici faire une revue détaillée de chacune des mentions du référent du pronom personnel *elle* comme nous l'avons fait jusqu'à présent. Nous nous contenterons donc d'une vue générale des compositions des CDR comportant 3 maillons (ou plus) + P1.

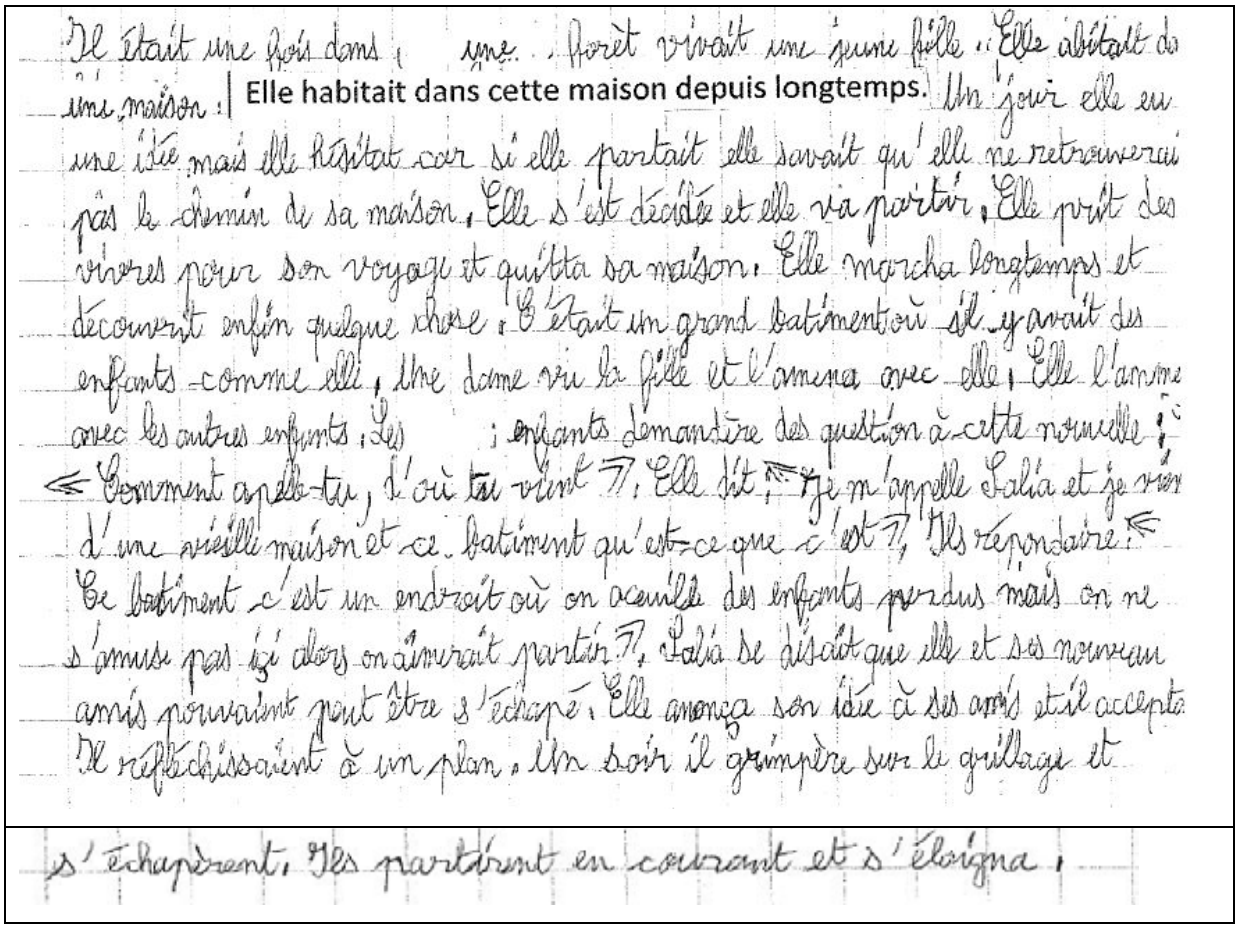
Quel que soit le nombre de mentions dans la CDR incluant le pronom personnel *elle*, le SN pronominal est la forme la plus fréquemment employée pour maintenir la référence. Ce résultat est en accord avec les différents travaux qui s'intéressent au maintien de la référence chez les enfants. En effet, Karmiloff-Smith (1980, 1981 ; cf. chapitre 3, section 3.1.2.2.) notamment, montre que les pronoms sont les formes privilégiées de la reprise, d'autant plus lorsque le sujet grammatical est thématisé. Il est donc normal que ce soit la forme la plus fréquente pour la troisième mention du référent du pronom personnel *elle*. Le pronom est la

forme « non marquée » de la reprise (Kleiber, 1994 ; cf. chapitre 2, section 2.2.4.), donc la plus facilement et longtemps utilisable, jusqu'à la nécessité d'une redénomination.

En ce qui concerne les expansions des mentions du référent, on trouve des formes expansées à l'aide d'un modifieur dans toutes les CDR comportant jusqu'à 8 mentions + P1, même s'il s'agit d'un ou deux textes dans chaque type de CDR.

Enfin, l'étude des fonctions syntaxiques des différentes mentions montre qu'elles sont le plus fréquemment en position de sujet syntaxique, comme pour le pronom personnel *elle* dans P1 et qui fait du personnage féminin le sujet thématique. Les élèves vont donc avoir tendance à le thématiser également dans les mentions antérieures. C'est la stratégie du sujet thématique (Karmiloff-Smith, 1980, 1981 ; cf. chapitre 3, partie 3.1.2.3.).

Nous retrouvons ces éléments dans le texte de Mathieu (CM1 T1), inachevé, que nous reproduisons dans son intégralité et qui comporte la plus longue CDR incluant le pronom personnel *elle* dans notre corpus, puisqu'elle comporte 20 mentions du référent + P1.


Il était une fois dans une forêt vivait une jeune fille. Elle habitait dans une maison. Elle habitait dans cette maison depuis longtemps. Un jour elle eu une idée mais elle hésitat car si elle partait elle savait qu'elle ne retrouverai pas le chemin de sa maison. Elle s'est décidée et elle va partir. Elle prit des vivres pour son voyage et quitta sa maison. Elle marcha longtemps et découvrit enfin quelque chose. C'était un grand batiment où il y avait des enfants comme elle. Une dame vu la fille et l'amena avec elle. Elle l'amene avec les autres enfants. Les enfants demandère des question à cette nouvelle : « Comment apelle-tu, d'où tu vient ? ». Elle dit : « Je m'appelle Salia et je vien d'une vieille maison et ce batiment qu'est-ce que c'est ? ». Ils répondaire : « Ce batiment c'est un endroit où on acuille des enfants perdus mais on ne s'amuse pas ici alors on aimerait partir ? ». Salia se disait que elle et ses nouveau amis pouvaient peut être s'échapé. Elle anonça son idée à ses amis et il accepta. Il réfléchissaient à un plan. Un soir il grimpère sur le grillage et s'échapèrent. Ils partèrent en courant et s'éloigna.

Il était une fois dans une forêt vivait une jeune fille. Elle habitait dans une maison. *Elle habitait dans cette maison depuis longtemps.* Un jour, elle eut une idée mais elle hésita car si elle partait, elle savait qu'elle ne retrouverait pas le chemin de sa maison. Elle s'est décidée et elle va partir. Elle prit des vivres pour son voyage et quitta sa maison. Elle marcha longtemps et découvrit enfin quelque chose. C'était un grand bâtiment où il y avait des enfants comme elle. Une dame vit la fille et l'amena avec elle. Elle l'amena avec les autres enfants. Les enfants demandèrent des questions à cette nouvelle : « Comment t'appelles-tu ?, d'où tu viens ? ». Elle dit : « Je m'appelle Salia et je viens d'une vieille maison et ce bâtiment, qu'est-ce que c'est ? ». Ils répondirent : « Ce bâtiment, c'est un endroit où on accueille des enfants perdus mais on ne s'amuse pas ici alors on aimerait partir ». Salia se disait qu'elle et ses nouveaux amis pouvaient peut-être s'échapper. Elle annonça son idée à ses amis et ils acceptèrent. Ils réfléchissaient à un plan. Un soir, ils grimpèrent sur le grillage et s'échappèrent. Ils partirent en courant et s'éloignèrent.

Figure 53 : Texte de Matthieu, CM1 T1

Dans ce texte, le référent du pronom personnel est introduit dans la première phrase par l'expression indéfinie *une jeune fille,* que nous appellerons expression indéfinie *collocative* car il s'agit d'une association (entre *jeune* et *fille*) très fréquemment observées dans les textes d'élèves. Dans la phrase suivante, étant donné que le référent est maintenant connu du lecteur et accessible, il est repris par le pronom personnel *elle*. Après P1, le référent est maintenu principalement à l'aide du pronom personnel *elle* en position de sujet syntaxique. On trouve toutefois deux formes atténuées de la référence, il s'agit de pronom zéro dans *et Ø quitta sa maison* et *et Ø découvrit enfin quelque chose*. L'emploi de ce type de forme témoigne d'un degré plus haut dans la maîtrise du maintien de la référence.

En outre, on constate également l'emploi du pronom *l'*en position de complément de verbe dans *l'amena avec elle*. Ce changement de position syntaxique coïncide avec le changement de sujet thématique : c'est le personnage d'*une dame* qui devient saillant. Celui de la jeune fille, moins saillant, est d'ailleurs d'abord repris par le SN défini *la fille* avant de l'être par le pronom *l'* puis par le SN démonstratif *cette nouvelle*. On remarque dans ce passage un effort de variété dans les formes lexicales utilisées.

Enfin, l'emploi du nom propre *Salia* dans la CDR montre également un degré de maîtrise plus élevé. En effet, le référent du pronom personnel *elle* est trop éloigné, sémantiquement parlant, de sa dernière mention dans le texte et nécessite donc d'être

réinstancié, conformément à la méta-règle de relation (Charolles, 1978 ; cf. chapitre 1, section 1.1.2). Les personnages des enfants, rendus saillants avant le dialogue et la diversité des informations contenues dans la conversation font que le référent ne peut pas être récupéré par le lecteur sous la forme d'un pronom. La réinstanciation s'opère donc sous la forme du nom propre *Salia*. La redénomination vient ici réactiver le référent en question.

8.3. Bilan de l'étude du référent du pronom personnel *elle*

De manière générale, l'étude des formes de la première mention du référent du pronom personnel *elle* montre que :

- la forme privilégiée pour introduire le référent est le SN indéfini, avec une fréquence d'emploi globalement croissante en fonction du niveau scolaire.
- la première mention du référent apparaît majoritairement en position de sujet syntaxique, mais également très fréquemment dans une construction présentative. Les élèves les plus jeunes ont plutôt tendance à utiliser la position de sujet alors que les élèves les plus âgés emploient plutôt des constructions présentatives.
- les formes lexicales des SN employés en première mention sont principalement des dénominations « passe-partout » : *fille, dame, femme…* Les noms fonctionnels (*sorcière, voisine…*) et les noms relationnels sont moins nombreux, quel que soit le niveau scolaire.
- lorsque la forme de la première mention est expansée, elle l'est principalement à l'aide d'une seule expansion. Toutefois, plus le niveau scolaire augmente, plus le nombre d'expansion augmente.
- les types d'expansion utilisés sont principalement des SN adjectivaux dans la première expansion et des propositions relatives dans la deuxième. Plus le niveau scolaire des élèves augmente, plus la fréquence d'emploi des SN adjectivaux et des propositions relatives augmente.

L'étude du maintien de la référence au pronom personnel *elle* montre plusieurs manières de « gérer » la référence. La première est une gestion plutôt locale du référent : les textes présentent deux mentions du référent, c'est-à-dire 1 mention + celle de P1. Dans

cette configuration, la mention employée par les élèves (ou source) est localisée soit avant P1 (anaphore), soit après P1 (cataphore). Selon la localisation de la source, la forme de la première mention change. Dans les cas d'anaphore, la forme la plus fréquente est le SN indéfini. Dans les cas de cataphore, c'est le pronom personnel *elle*. Cependant, quelle que soit la localisation de la source, la fonction syntaxique la plus fréquente est le sujet.

La deuxième manière de « gérer » la référence se traduit par une gestion plus globale du référent : les élèves le mentionnent plusieurs fois dans leur texte, ce qui constitue des CDR plus ou moins longues. Lorsque la CDR déployée pour le pronom personnel *elle* est minimale (2 mentions + P1, soit 3 mentions au total), ou bien plus longue (3 mentions ou plus + P1), on constate une prédominance de SN pronominaux comme formes de reprise du référent, en position de sujet syntaxique. On constate également que plus le niveau scolaire augmente, plus la longueur des CDR augmente.

CHAPITRE 9 : Introduction et maintien du référent du pronom personnel *il*

Dans la tâche-problème que doivent réaliser les élèves de notre étude, la deuxième phrase à insérer (P2) est la suivante :

Il se retourna en entendant ce grand bruit.

Dans cette phrase, le pronom personnel *il* nécessite au préalable l'introduction d'un référent car l'emploi d'un pronom implique un degré d'accessibilité fort de l'entité à laquelle il réfère (cf. chapitre 1, section 1.2.2.). En effet, le pronom ne comportant pas d'informations sur le référent, il faut la présence d'une source pour établir la référence. Dans P2, le pronom personnel *il* est marqué en genre. Toutes les expressions coréférentielles devront alors respecter la marque du masculin. En revanche, aucune indication n'est fournie quant à la nature humaine ou non, animée ou non de l'entité.

En 9.1., nous étudions la manière dont le référent du pronom personnel *il* est introduit, en analysant à la fois les formes de la première mention du référent, leurs éventuelles expansions et leurs fonctions syntaxiques. Après avoir été introduit, le référent peut être maintenu tout au long du discours, ou tout du moins dans une partie du discours. En 9.2., nous nous intéressons, de ce fait, à la manière dont la référence est maintenue tout au long du texte.

9.1. Introduction du référent du pronom personnel *il*

De manière générale, dans notre corpus, nous observons trois localisations possibles de l'expression référentielle qui introduit le référent (ou *source*), comme le montre le tableau 14 :

- La source est introduite avant la phrase imposée et la référence est maintenue après
- La source n'est pas introduite avant la phrase imposée mais elle est localisée après
- La source est absente : la seule mention du référent est celle de la phrase imposée

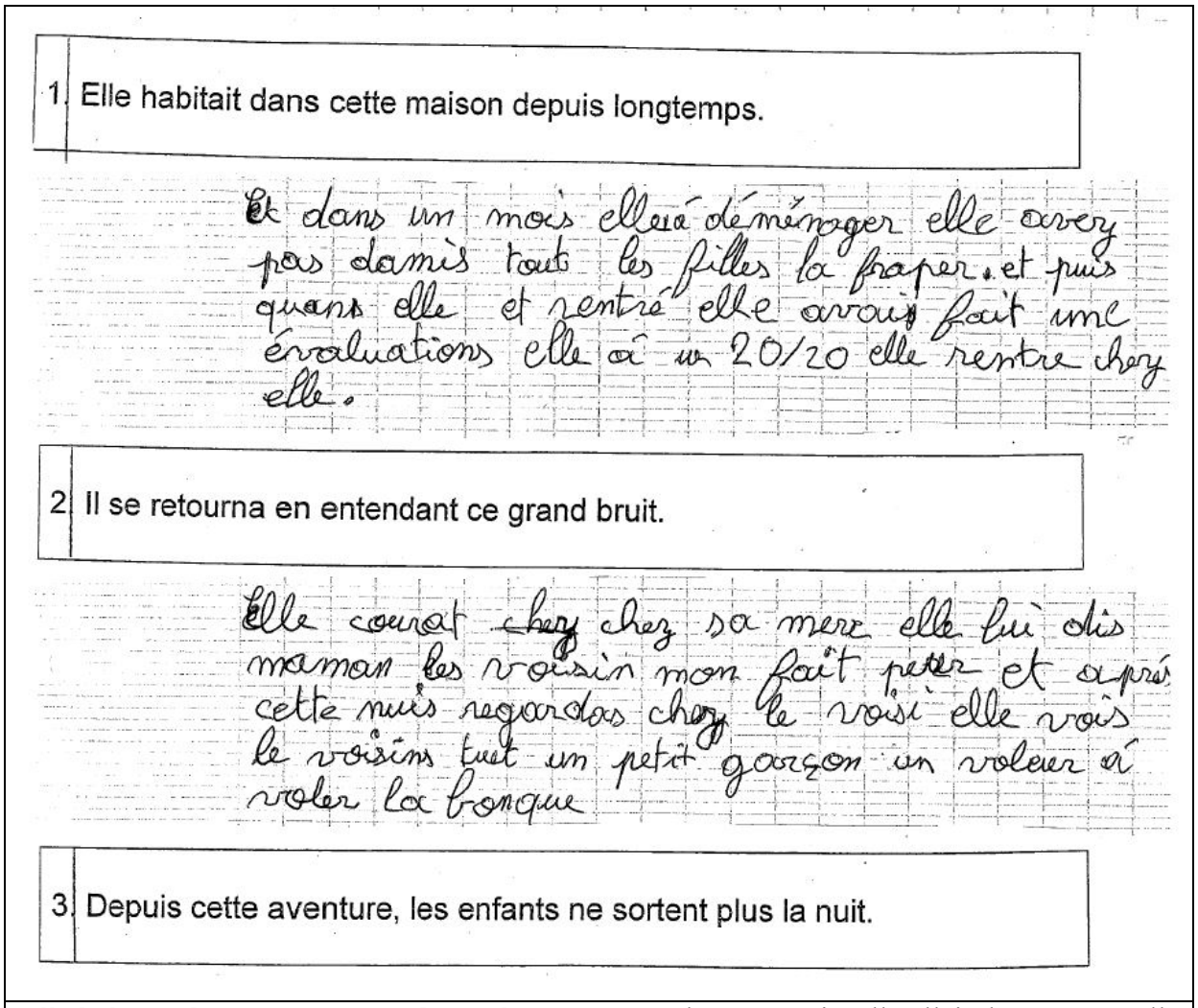
Localisation de la source du pronom personnel *il*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
avant P2 et maintien après	53	**83%**	46	**82%**	59	**91%**	158	**85%**
après P2	5	8%	0	0%	2	3%	7	**4%**
source absente	6	9%	10	18%	4	6%	20	11%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 14 : Localisation de la source du référent du pronom personnel *il*

Sur l'ensemble du corpus, le référent est introduit avant P2 dans 85% des textes et maintenu par la suite dans le discours. Globalement, plus le niveau scolaire augmente, plus les textes dans ce cas sont nombreux : 83% au CE2, 82% au CM1, 91% au CM2. En effet, les élèves avançant en âge et niveau scolaire, ils sont plus nombreux à savoir gérer à la fois l'introduction et le maintien de la référence (cf. chapitre 3, section 3.1.).

De manière marginale, la source est localisée après P2 dans 4% des textes du corpus (7 textes). Il s'agit de cataphores (cf. chapitre 2, section 2.1.1.1.) : le pronom personnel *il* de P2 est la première mention du référent. Cette manière d'introduire le référent semble plus fréquente lorsque les élèves sont plus jeunes : entre 8% au CE2 et 3% au CM2. On constate toutefois qu'au CM1, aucun des textes ne présente de cataphore.

Enfin, sur l'ensemble du corpus, dans 11% des textes, la source est absente. C'est-à-dire que la seule mention du référent du pronom personnel *il* est celle contenue dans P2, comme dans (54).



1. Elle habitait dans cette maison depuis longtemps.

Et dans un mois elleá déménager elle avey pas damis tout les filles la fraper. et puis quans elle et rentré elle avais fait une évaluations elle aí un 20/20 elle rentre chez elle.

2. Il se retourna en entendant ce grand bruit.

Elle courat ~~chez~~ chez sa mere elle lui dis maman les voisin mon fait peter et aprés cette nuis regardas chez le voisi elle vois le voisins tuet un petit garçon un voleur aí voler la banque

3. Depuis cette aventure, les enfants ne sortent plus la nuit.

Elle habitait dans cette maison depuis longtemps. Et dans un mois, elle allait déménager. Elle n'avait pas d'ami, toutes les filles la frappaient. Et puis quand elle est rentrée, elle avait fait une évaluation. Elle a eu 20/20. Elle rentre chez elle. *Il se retourna en entendant ce grand bruit.* Elle courut chez sa mère. Elle lui dit : « maman, le voisin m'a fait peur ». Et après cette nuit regarda chez le voisin. Elle voit le voisin tuer un petit garçon. Un voleur a volé la banque. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 54 : Texte de Rayene, CM1 B

On pourrait s'attendre à ce que la fréquence de l'absence de source décroisse avec l'âge. Or, malgré le fait que notre corpus est limité, on constate tout de même que c'est au CM1 que les textes dans lesquels il n'y a aucune source sont les plus nombreux : 18% au CM1 contre 9% au CE2 et 6% au CM2. Toutefois, le nombre de textes concernés étant faible, nous nous garderons de toute généralisation.

Après avoir donné une vue d'ensemble des différentes manières de gérer la référence au personnage, nous allons étudier les formes de la première mention du référent du pronom personnel *il.*

9.1.1. Etude des formes de la première mention du référent du pronom personnel *il*

Nous venons de le voir, le pronom personnel *il* nécessite une source introduite préalablement à P2. Cette source constitue la première mention (M1) du référent du pronom personnel *il*. Le tableau 15 reproduit les différents types de formes de la première mention du référent du pronom personnel *il.*

Formes de M1 pour *il*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN indéfini	16	25%	15	27%	26	40%	57	**31%**
SN possessif	16	25%	11	20%	14	22%	41	**22%**
SN nom propre	9	14%	10	18%	7	11%	26	14%
SN défini	8	13%	5	9%	9	14%	22	12%
SN numéral	2	3%	2	4%	1	2%	5	3%
SN pronominal	7	11%	1	2%	4	6%	12	6%
SN démonstratif	0	0%	1	2%	0	0%	1	1%
P2 = seule mention	6	9%	11	20%	4	6%	21	11%
Total	64	100%	56	100%	65	100%	185	100%

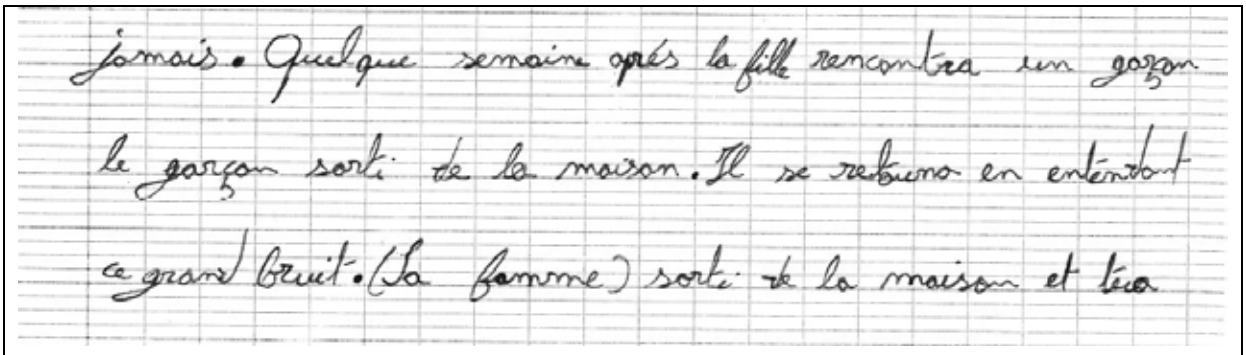
Tableau 15 : Formes de la première mention du référent du pronom personnel il

***Les différentes formes d'introduction du référent du pronom personnel* il**

Les différentes études qui s'intéressent à l'introduction des référents dans un discours chez les enfants montrent que la forme d'introduction privilégiée est le SN indéfini. On peut citer par exemple les travaux de Hickmann, Kail & Roland (1995 ; cf. chapitre 3, section 3.1.3.) qui étudient comment la connaissance partagée ou non affecte l'évolution de l'utilisation des expressions référentielles par des enfants de 6, 9 et 11 ans. L'étude révèle qu'à partir de 9 ans, les enfants utilisent principalement des formes linguistiques indéfinies

pour introduire les référents en situation de connaissance non partagée. Ces travaux indiquent aussi que c'est à 11 ans (CM2) que les enfants emploient plus systématiquement les formes indéfinies pour l'introduction des référents, quel que soit le type de situation (connaissance partagée ou non). La situation de connaissance non partagée s'apparente également à la production d'écrit. En effet, lors de la lecture d'un texte, le lecteur/récepteur ne peut s'aider d'un quelconque appui visuel pour décoder le message et l'encodeur/scripteur ne peut miser sur le fait que le lecteur partage les mêmes informations que lui. Il fait donc en sorte de donner au lecteur toutes les informations nécessaires à la construction de la référence.

Nous retrouvons ces résultats dans nos propres travaux. En effet, la première mention est principalement introduite sous la forme d'un SN indéfini : 31% sur l'ensemble du corpus, comme l'illustre (55) où le pronom personnel *il* de P2 réfère au SN indéfini *un garçon* :


jamais. Quelque semaine après la fille rencontra un garçon
le garçon sorti de la maison. Il se retourna en entendant
ce grand bruit. (La femme) sorti de la maison et lui

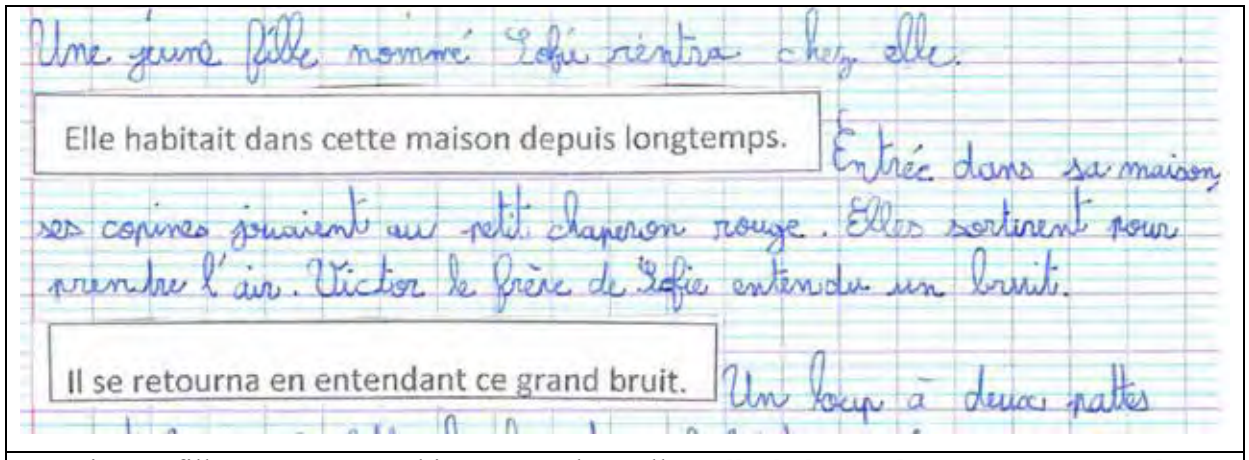
Quelques semaines après, la fille rencontra un garçon. Le garçon sortit de la maison. *Il se retourna en entendant ce grand bruit.*

Figure 55 : Texte de Mathias, CM2 M

La fréquence d'emploi d'un SN indéfini en première mention du référent n'est pas identique d'une classe à l'autre. En effet, si elle croît légèrement entre le CE2 (25%) et le CM1 (27%), on constate une augmentation significative au CM2 (40%). Cet écart, déjà constaté dans les travaux précédemment cités, peut s'expliquer par le fait que le recours au SN indéfini est considéré comme une stratégie experte. En effet, les expressions nominales indéfinies sont autonomes référentiellement parlant car elles ne demandent pas à ce que l'interlocuteur sache déjà de quoi il va être question dans le propos (Charolles, 2002 ; cf. chapitre 1, section 1.4.2.1). Les SN indéfinis sont donc utilisés lorsque le locuteur introduit pour la première fois l'objet du discours (cf. chapitre 2, section 2.1.1.3.), ce qui est le cas ici, ou un nouveau référent. Cette stratégie d'introduction se retrouve chez les scripteurs experts.

Corblin (1985) ou Longo et Todirascu (2009) ont étudié et comparé la nature des chaînes de référence dans plusieurs genres, et notamment le genre narratif, tel que les romans d'auteurs reconnus. Il s'avère que dans ce genre, les premiers maillons des chaînes de référence sont principalement des SN indéfinis. Le fait de retrouver ces résultats chez nos apprentis scripteurs les plus âgés témoigne donc d'un degré de maîtrise dans l'introduction du référent plus élevé.

Pour introduire le référent du pronom personnel *il*, on trouve également des noms propres. Ce type de forme est la troisième forme la plus fréquente : 14% sur l'ensemble du corpus, comme l'emploi de *Victor* dans l'exemple (56) :

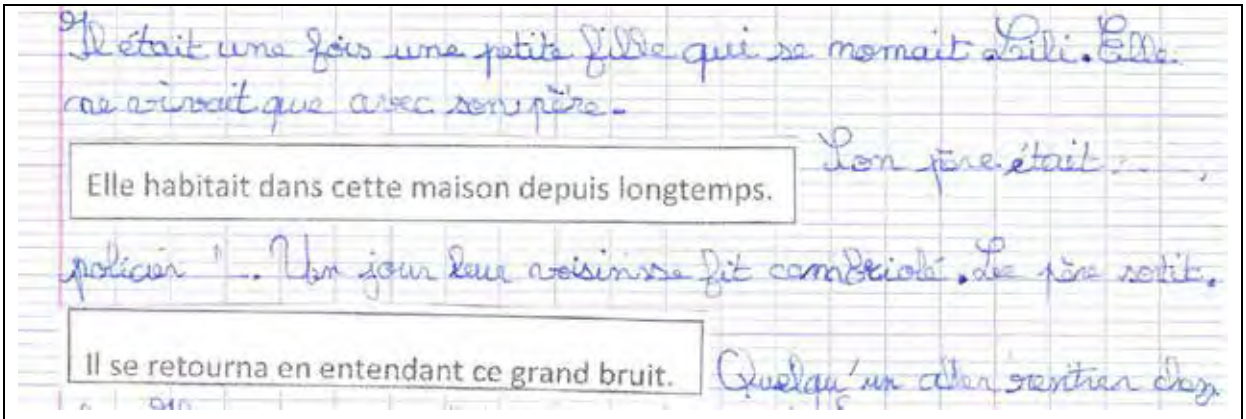


Une jeune fille nommée Sophie rentra chez elle. *Elle habitait dans cette maison depuis longtemps.* Entrée dans sa maison, ses copines jouaient au petit chaperon rouge. Elles sortirent pour prendre l'air. Victor, le frère de Sophie, entendit un bruit. *Il se retourna en entendant ce grand bruit*.

Figure 56 : Texte de Thomas, CM1 T2

De même que le SN indéfini, le nom propre permet d'introduire l'objet du discours. En effet, étant donné qu'il réfère à un particulier en le discriminant au sein d'une catégorie d'autres particuliers, le nom propre réfère « à des particuliers qui ne sont pas supposés déjà présents dans l'esprit des interlocuteurs » (Charolles, 2002 :74), c'est-à-dire lorsque l'accessibilité du référent est faible (Ariel, 1990 ; cf. chapitre 1, section 1.2.2.). Par cette opération, le nom propre porte l'attention sur une entité non encore saillante et signale qu'elle a de fortes chances de le devenir par la suite.

L'analyse de notre corpus révèle un fait surprenant : la deuxième forme la plus fréquente pour introduire le référent du pronom personnel *il* est le SN possessif (22% pour l'ensemble du corpus), comme l'illustre l'exemple (57).

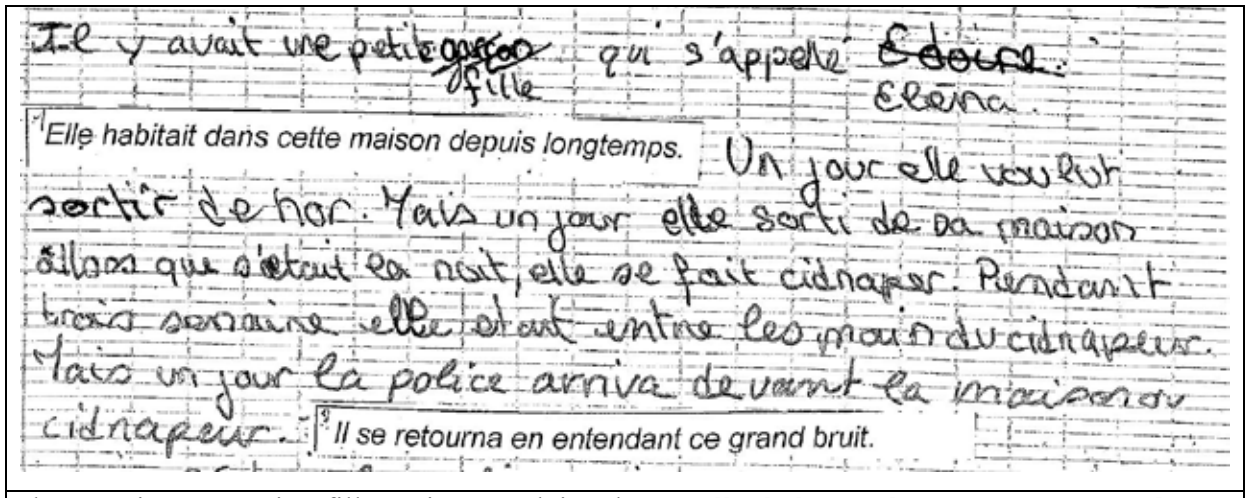


Il était une fois une petite fille qui se nommait Lili. Elle ne vivait qu'avec son père. *Elle habitait dans cette maison depuis longtemps.* Son père était policier. Un jour, leur voisin se fit cambrioler. Le père sortit. *Il se retourna en entendant ce grand bruit.*

Figure 57 : Texte de Marine, CE2 T

Le SN possessif est la deuxième forme la plus répandue sur l'ensemble du corpus, mais ce n'est pas le cas pour le niveau du CE2. Dans les productions de ce niveau, il y autant de premières mentions du référent sous la forme d'un SN indéfini que d'un SN possessif : 25% pour les deux. Cela peut tenir au fait qu'il est plus facile pour les élèves les plus jeunes d'introduire un nouveau référent en lien avec un référent déjà introduit précédemment. En d'autres termes, il semble plus facile d'amener un nouveau personnage grâce à un lien de parenté par exemple avec un autre personnage, plutôt que d'introduire un personnage sans lien aucun avec les autres.

Enfin, nous remarquons l'emploi de SN définis en première mention du référent du pronom personnel *il* mais il s'agit principalement de descriptions définies complètes. Les descriptions définies complètes ne codent qu'une seule entité ou un seul groupe d'entités et permettent ainsi de les identifier indépendamment de toute information sur le contexte où elles sont employées (Charolles, 2002 ; cf. chapitre 1, section 1.4.2.1.). Dans les textes de notre corpus, il s'agit principalement de SN définis qui introduisent le personnage à l'aide d'une description définie complète permettant de l'identifier comme étant le seul référent possible (*le garçon qui s'appelle Ali*) ou de SN définis qui introduisent le personnage par sa fonction (*le marchand, le chef, etc.),* comme *le kidnappeur* dans (58) :


Il y avait une petite ~~garçon~~ fille qui s'appelé ~~Edoure~~ Eléna.
Elle habitait dans cette maison depuis longtemps. Un jour elle voulut sortir de hor. Mais un jour elle sorti de sa maison alors que c'était la nuit, elle se fait cidnaper. Pendant trois semaine elle était entre les main du cidnapeur. Mais un jour la police arriva devant la maison du cidnapeur. Il se retourna en entendant ce grand bruit.

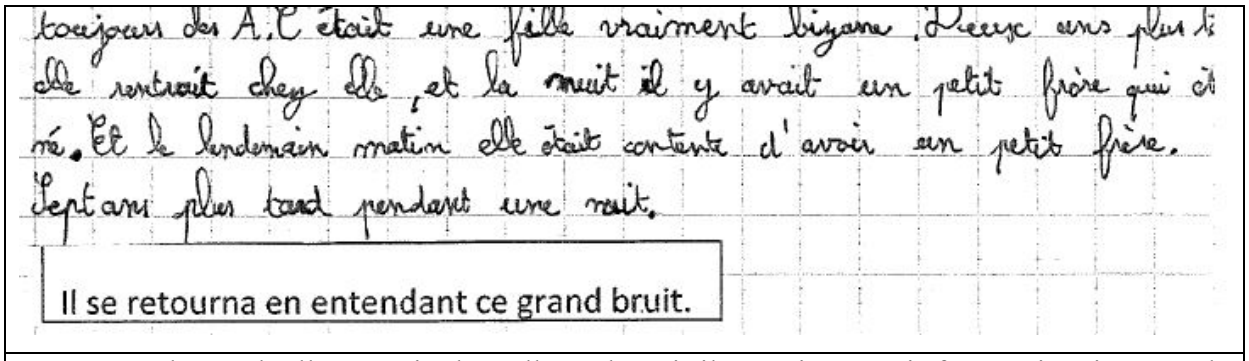
Il y avait une petite fille qui s'appelait Eléna. *Elle habitait dans cette maison depuis longtemps.* Un jour, elle voulut sortir dehors. Mais un jour, elle sortit de sa maison alors que c'était la nuit : elle se fait kidnapper. Pendant trois semaines, elle était entre les mains du kidnappeur. Mais un jour, la police arriva devant la maison du kidnappeur. *Il se retourna en entendant ce grand bruit.*

Figure 58 : Texte d'Antoine, CM2 B

Étude des formes lexicales de M1 pour il

La diversité des formes lexicales employées pour référer au pronom personnel *il* montre différentes manières de « concevoir » le personnage. Pour en rendre compte, nous nous inspirons de catégories empruntées à Charolles (1988c ; cf. chapitre 3, section 3.1.2.3), auxquelles nous ajoutons les noms relationnels. Ainsi, la référence au personnage masculin peut se réaliser de diverses manières : soit par des dénominations « passe-partout » faisant appel à l'identité du personnage masculin (*garçon, enfant, homme, monsieur…)*, soit par des noms relationnels au sein du noyau familial (*fils, petit-frère, grand-père…*), soit par des dénominations faisant appel au rôle du personnage dans l'histoire que nous appelons *noms fonctionnels* (*marchand, voisin, policier…).* On trouve également deux textes dans lesquels le référent du pronom personnel *il* est une entité non humaine : il s'agit d'un chien et d'un dinosaure

L'étude des formes lexicales de la première mention du référent du pronom personnel *il* monte que les noms relationnels sont les plus fréquemment employés pour introduire le personnage masculin, quel que soit le niveau de classe : 43% dans l'ensemble du corpus, comme l'illustre (59).


toujours des A.C était une fille vraiment bizarre. Deux ans plus ti
elle rentrait chez elle, et la nuit il y avait un petit frère qui est
né. Et le lendemain matin elle était contente d'avoir un petit frère.
Sept ans plus tard pendant une nuit,

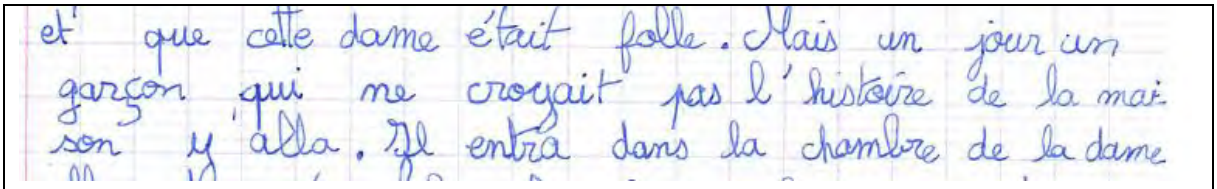
Il se retourna en entendant ce grand bruit.

Deux ans plus tard, elle rentrait chez elle, et la nuit il y avait un petit-frère qui était né. Et le lendemain matin, elle était contente d'avoir un petit frère. Sept ans plus tard, pendant une nuit, *il se retourna en entendant ce grand bruit.*

Figure 59 : Texte de Corentin, CM1 T1

Tous les noms masculins du noyau familial sont représentés : *mari, père, fils, grand-père, petit-fils, petit-frère, grand-frère*, bien que la forme *frère* soit majoritaire. On les retrouve plus fréquemment lorsque la première mention du référent est un SN possessif.

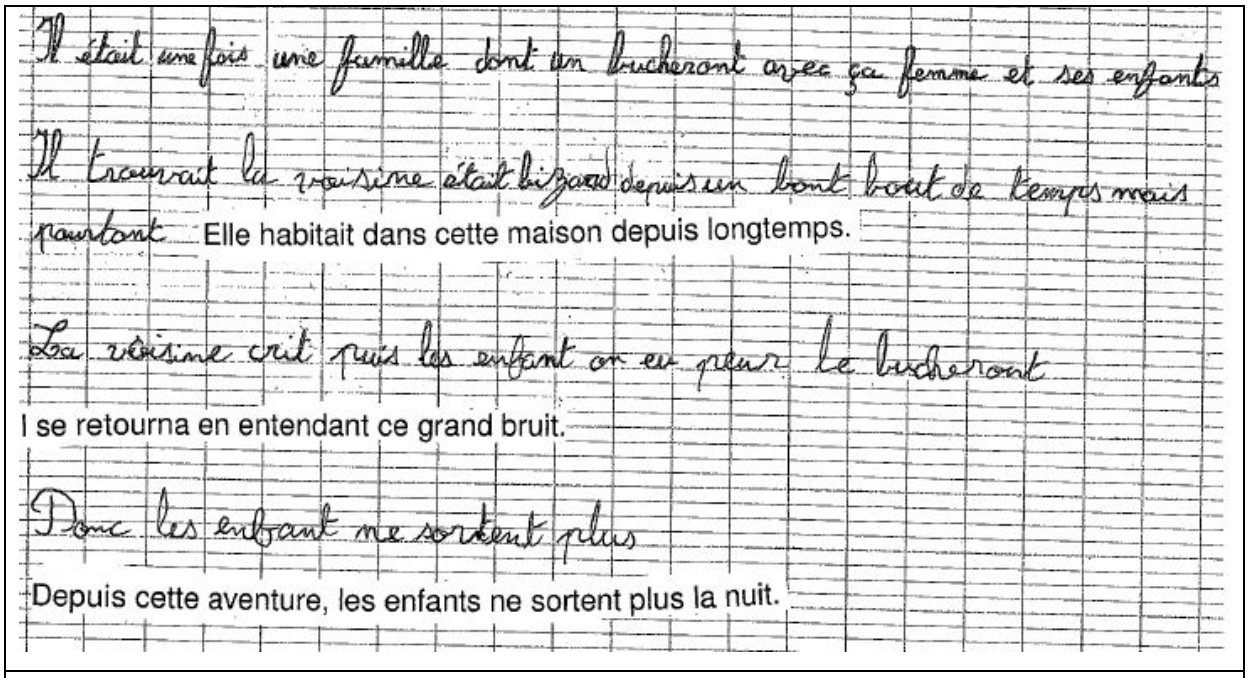
Après les noms relationnels, les noms génériques sont les plus fréquemment employés pour la première mention du référent du pronom personnel *il* : 36% des textes. Il s'agit majoritairement de la forme lexicale *garçon*, comme dans (60) :


et que cette dame était folle. Mais un jour un
garçon qui ne croyait pas l'histoire de la mai-
son y alla. Il entra dans la chambre de la dame

Mais un jour, un garçon qui ne croyait pas l'histoire de la maison y alla. Il entra dans la chambre de la dame.

Figure 60 : Texte de Léonore, CE2 T

Enfin, les noms fonctionnels sont les moins fréquents : 21% dans l'ensemble du corpus, comme dans (61).

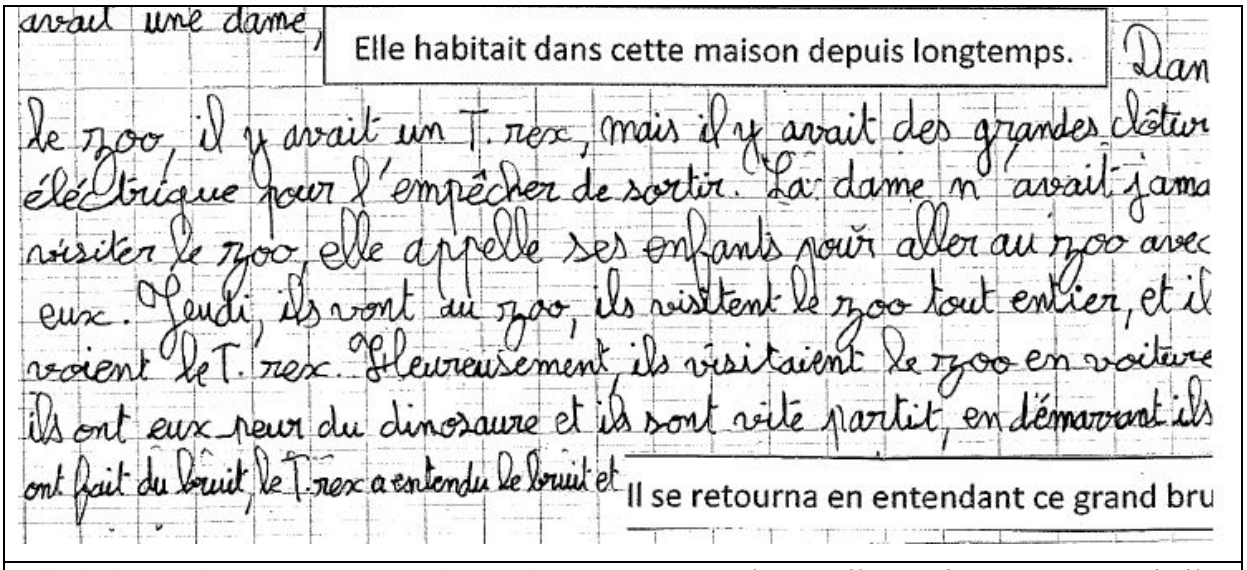

Il était une fois une famille dont un bucheront avec ça femme et ses enfants
Il trouvait la voisine était bizard depuis un bont bout de temps mais pourtant Elle habitait dans cette maison depuis longtemps.
La vôisine crit puis les enfant ont eu peur le bucheront
I se retourna en entendant ce grand bruit.
Donc les enfant ne sortent plus
Depuis cette aventure, les enfants ne sortent plus la nuit.

Il était une fois une famille dont un bûcheron avec sa femme et ses enfants. Il trouvait la voisine était bizarre depuis un bout de temps mais pourtant, *elle habitait dans cette maison depuis longtemps.* La voisine crie puis les enfants ont eu peur. Le bûcheron, *il se retourna en entendant ce grand bruit.* Donc les enfants ne sortent plus. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 61 : Texte de Néal, CM2 S

Les noms fonctionnels apparaissent plutôt chez les élèves les plus âgés mais restent minoritaires dans l'ensemble du corpus.

Enfin, nous avons également constaté deux formes lexicales représentant une entité non humaine. Il s'agit d'un texte dans lequel le référent du pronom personnel *il* est un chien et d'un texte dans lequel c'est un dinosaure, comme l'illustre l'exemple (62) :


avait une dame, Elle habitait dans cette maison depuis longtemps. Dan
le zoo, il y avait un T. rex, mais il y avait des grandes clôtur
électrique pour l'empêcher de sortir. La dame n'avait jama
visiter le zoo, elle appelle ses enfants pour aller au zoo avec
eux. Jeudi, ils vont au zoo, ils visitent le zoo tout entier, et il
voient le T. rex. Heureusement, ils visitaient le zoo en voiture
ils ont eux peur du dinosaure et ils sont vite partit, en démarrant ils
ont fait du bruit, le T. rex a entendu le bruit et Il se retourna en entendant ce grand bru

Elle habitait dans cette maison depuis longtemps. Dans le zoo, il y avait un T-Rex, mais il y avait des grandes clôtures électriques pour l'empêcher de sortir. La dame n'avait jamais visité le zoo, elle appela ses enfants pour aller au zoo avec eux. Jeudi, ils vont au zoo, ils visitent le zoo tout entier, et ils voient le T-Rex. Heureusement, ils visitaient le zoo en voiture. Ils ont eu peur du dinosaure et ils sont vite partis, en démarrant ils ont fait du bruit, le T-Rex a entendu le bruit et *il se retourna en entendant ce grand bruit.*

Figure 62 : Texte d'Alix, CM1 T1

9.1.2. Etude des expansions de la première mention du référent du pronom personnel *il*

Lors de l'introduction du référent du pronom personnel *il* présent dans P2, on trouve dans les textes des SN nus et des SN expansés, présentant un ou plusieurs modifieurs qui sont autant d'ajouts d'informations à propos de l'entité ainsi dénotée (Charolles, 2002).

L'étude des formes de la première mention du référent du pronom personnel *il* montre que si, dans leur majorité, les textes du corpus ne présentent pas d'expansion, ceux qui en présentent une seule (35%) sont plus nombreux que ceux qui en présentent deux (2%, soit 3 textes), comme le montre le tableau 16 suivant.

Nbre d'expansions de M1 pour *il*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
1 expansion	22	34%	20	36%	22	34%	64	**35%**
2 expansions	0	0%	1	2%	2	3%	3	**2%**
pas d'expansion	42	66%	35	63%	41	63%	118	64%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 16 : Nombre d'expansions dans les formes de la première mention du référent du pronom personnel *il*

Étude des types de première expansion dans la première mention du référent du pronom personnel il

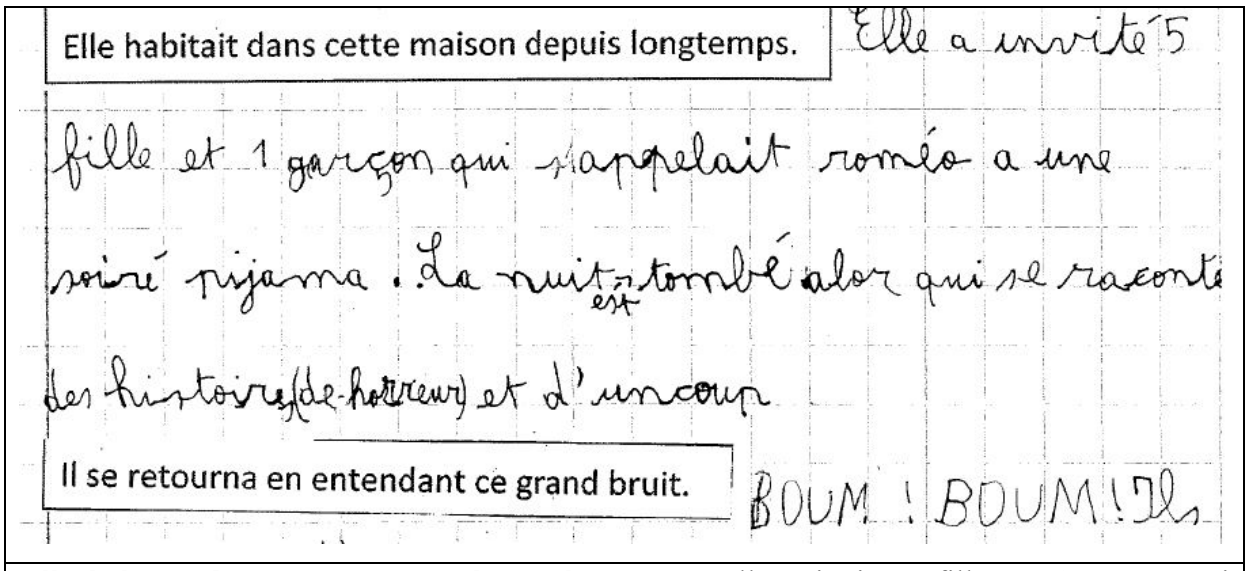
Les types d'expansion utilisés peuvent être de différentes sortes : syntagme adjectival, proposition relative, syntagme prépositionnel ou nom propre, comme on le voit dans le tableau 17.

Type de première expansion de M1 pour *il*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
syntagme adjectival	4	18%	5	24%	8	33%	17	25%
proposition relative	9	41%	8	38%	5	21%	22	33%
syntagme prépositionnel	2	9%	2	10%	2	8%	6	9%
autre	0	0%	2	10%	2	8%	4	6%
SN nom propre	7	32%	4	19%	7	29%	18	27%
Total	22	100%	21	100%	24	100%	67	100%

Tableau 17 : Type de première expansion dans les formes de la première mention du référent du pronom personnel *il*

L'étude des types d'expansions de la première mention du référent du pronom personnel *il* montre que lorsque le SN introduisant la première mention est expansé, il l'est le plus fréquemment à l'aide d'une proposition relative : 33% sur l'ensemble du corpus. La composition des propositions relatives employées est de deux sortes :

- des propositions relatives de dénomination, c'est-à-dire comportant un nom propre, caractérisées par la présence du verbe *s'appeler*, comme l'illustre l'exemple suivant avec la proposition relative *qui s'appelait Roméo* :


Elle habitait dans cette maison depuis longtemps. Elle a invité 5 fille et 1 garçon qui s'appelait roméo a une soiré pijama. La nuit est tombé alor qui se racontè des histoires (de horreur) et d'un coup

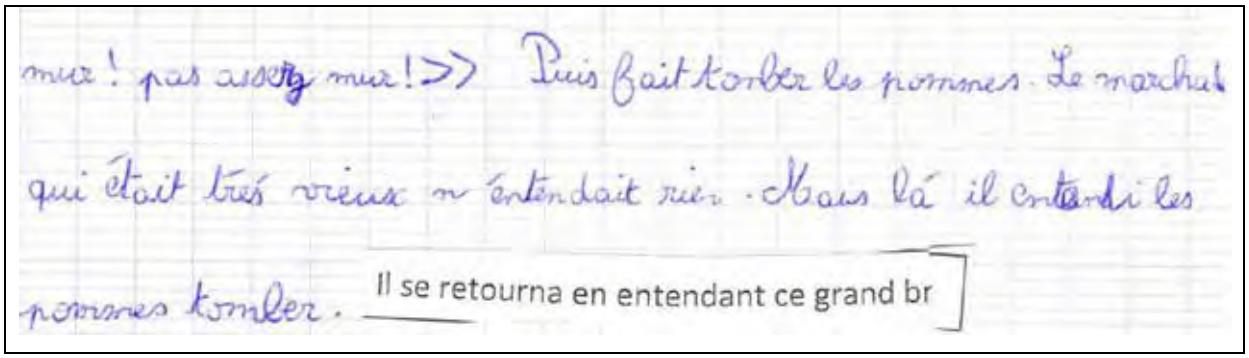
Il se retourna en entendant ce grand bruit. BOUM ! BOUM ! Ils

Elle habitait dans cette maison depuis longtemps. Elle a invité 5 filles et 1 garçon qui s'appelait Roméo à une soirée pyjama. La nuit est tombée alors qu'ils se racontaient des histoires (d'horreur) et d'un coup, *il se retourna en entendant ce grand bruit.*

Figure 63 : Texte de Mathéo, CM1 T1

L'utilisation d'une proposition relative comportant un nom propre est un atout dans la construction de la référence. En effet, ainsi introduit, le nom de baptême du personnage pourra être réutilisé dans la suite du texte, lorsqu'il sera nécessaire de réinstancier le personnage.

- des propositions relatives apportant des informations sur le personnage, comme par exemple le fait qu'il est vieux dans l'extrait (64) :


mur ! pas assez mur ! >> Puis fait tomber les pommes. Le marchand qui était très vieux n'entendait rien. Mais là il entendi les pommes tomber.

Il se retourna en entendant ce grand br

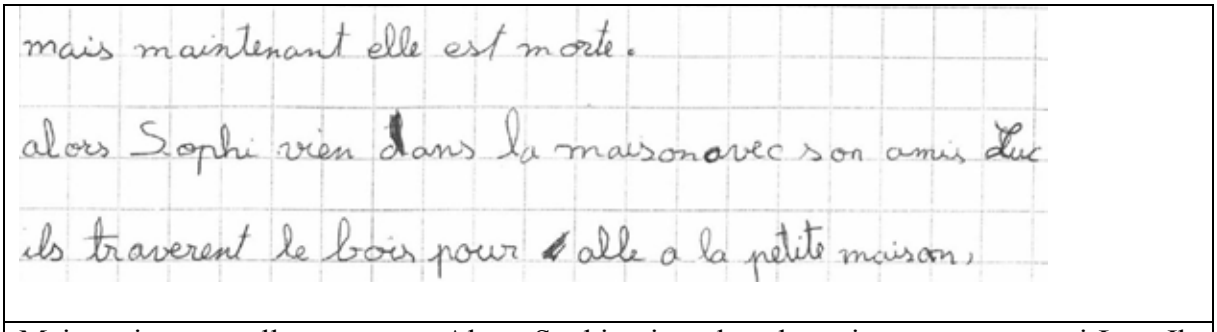
Puis fait tomber les pommes. Le marchand qui était très vieux n'entendit rien. Mais là, il entendit les pommes tomber.

Figure 64 : Texte de Naïla, CE2 P

Ce type d'information introduit à l'aide d'une proposition relative n'est pas utile à la référence au personnage, mais il apporte des précisions sur celui-ci et permet de rendre le texte plus riche.

Mises à part les propositions relatives, la première expansion du SN référant au personnage masculin peut être :

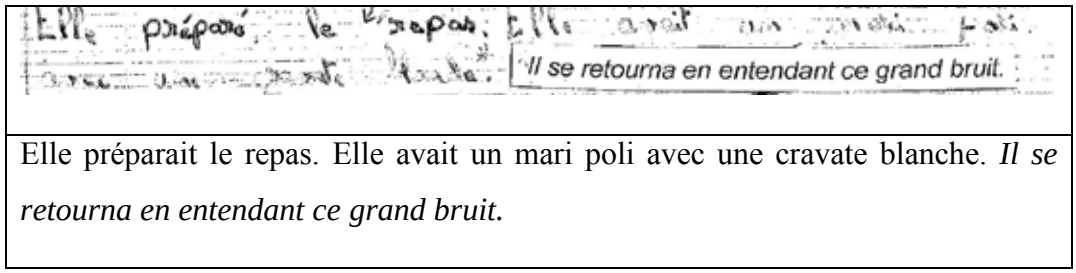
- un nom propre : 27% sur l'ensemble du corpus. On peut trouver ce type d'expansion sous la forme *nommé Matéo* ou directement un nom propre, comme c'est le cas dans (65) ci-dessous avec le SN possessif expansé *son ami Luc* :


mais maintenant elle est morte.
alors Sophi vien dans la maison avec son amis Luc
ils traverent le bois pour alle a la petite maison,

Mais maintenant elle est morte. Alors, Sophie vient dans la maison avec son ami Luc. Ils traversent le bois pour aller à la petite maison.

Figure 65 : Texte d'Aliénor, CE2 C

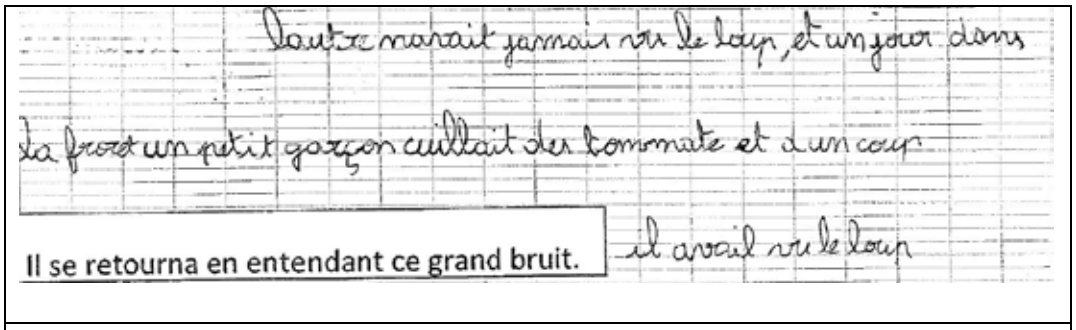
- un syntagme adjectival : 25% sur l'ensemble du corpus. Dans les syntagmes adjectivaux utilisés, nous distinguons deux formes. D'une part, il y a des adjectifs qualificatifs tels que *poli,* dans (66).


Il se retourna en entendant ce grand bruit.

Elle préparait le repas. Elle avait un mari poli avec une cravate blanche. *Il se retourna en entendant ce grand bruit.*

Figure 66 : Texte de Bastien, CM2 B

D'autre part, on remarque des associations privilégiées, telles que *petit garçon* dans (67), *jeune garçon* ou encore *jeune homme* :



L'autre n'avait jamais vu le loup, et un jour dans la forêt, un petit garçon cueillait des tomates et d'un coup, *il* se *retourna en entendant ce grand bruit.*

Figure 67 : Texte de Soliana, CM1 T1

Ces associations privilégiées de formes lexicales, ou collocations (Bally, 1909 ; Tutin & Grossman, 2002 ; cf. chapitre 1, section 1.3.4.), sont fréquentes dans la littérature de jeunesse. Par la fréquentation quotidienne de ce type de littérature, les collocations citées *supra* peuvent se retrouver dans les écrits des élèves.

- Un syntagme prépositionnel : 6 textes sur l'ensemble du corpus. Il s'agit de SN expansés tels que *son petit-frère de 3 ans*.

Dans les types de première expansion du SN référant au personnage masculin, on remarque donc que les propositions relatives sont les plus employées sur l'ensemble du corpus. Toutefois, la fréquence d'emploi des propositions relatives décroît avec le niveau scolaire : 41% au CE2, 38% au CM1, 21% au CM2. A l'inverse, la fréquence des syntagmes adjectivaux croît avec le niveau scolaire : 18% au CE2, 24% au CM1 et 33% au CM2. Cette différence témoigne d'un degré de maîtrise différent dans la référence. En effet, un élève qui ajoute une proposition relative à un SN, par exemple *un homme qui était vieux*, n'a pas le même degré de maîtrise de la référence qu'un élève qui intègre la même information au moyen d'un adjectif qualificatif, par exemple *un homme vieux.* Dans le premier cas, la proposition relative est caractéristique du langage oral des jeunes enfants. Plus les élèves avancent en âge et en niveau scolaire, moins leur discours écrit présente de caractéristiques du langage oral.

Comme nous l'avons vu dans le tableau 17 *supra,* le référent du pronom personnel *il* est majoritairement expansé à l'aide d'un seul modifieur. Dans le corpus, il n'y a que trois

textes dans lesquels le SN référant au pronom personnel *il* est expansé à l'aide de deux modifieurs : 1 texte au CM1 et 2 textes au CM2. Les types de deuxièmes expansions sont variés :

- une proposition relative qui apporte une information supplémentaire au personnage masculin : *qui n'avait que deux ans de moins que sa sœur* (CM1)
- un SN prépositionnel : *avec une cravate blanche* (CM2, cf. 67 *supra*)
- une expansion incluant un nom propre : *nommé Link* (CM2)

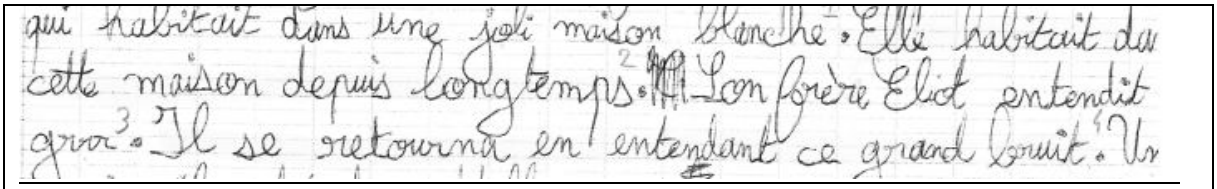
9.1.3. Étude des fonctions syntaxiques des formes de M1

Dans leur texte, les élèves sont amenés à introduire le référent du pronom personnel *il*. Lorsque la première mention de ce référent apparaît, c'est une information totalement nouvelle (cf. *statut de l'information*, chapitre 1, section 1.2.). Cette information représente également le thème de l'énoncé (cf. *progression thématique*, chapitre 1, section1.2.3.), qui peut se réaliser sous la forme de sujet syntaxique et se retrouver en position initiale dans la proposition, selon la stratégie du sujet thématique (Karmiloff-Smith, 1979).

Fonctions syntaxiques de M1 pour *il*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
sujet	36	56%	23	41%	35	54%	94	**51%**
construction présentative	2	3%	5	9%	3	5%	10	5%
phrase nominale	5	8%	3	5%	1	2%	9	5%
complément du nom	1	2%	0	0%	2	3%	3	2%
complément de verbe	13	20%	13	23%	19	29%	45	24%
complément de phrase	0	0%	1	2%	1	2%	2	1%
P2 = seule mention	7	11%	11	20%	4	6%	22	12%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 18 : Différents types de fonctions syntaxiques prises par le SN de la première mention du référent du pronom personnel *il*

En accord avec ces principes, l'étude des fonctions syntaxiques des formes de la première mention du référent du pronom personnel *il* montre que la fonction sujet est la plus fréquente dans les textes, quel que soit le niveau de classe : 51% sur l'ensemble du corpus, comme l'indique le tableau 18. Il s'agit de textes dans lesquels le référent du pronom personnel *il* est le sujet syntaxique, comme dans (68) où le SN possessif expansé *son frère Eliot*, introduit pour la première fois, est le sujet du verbe *entendre* :

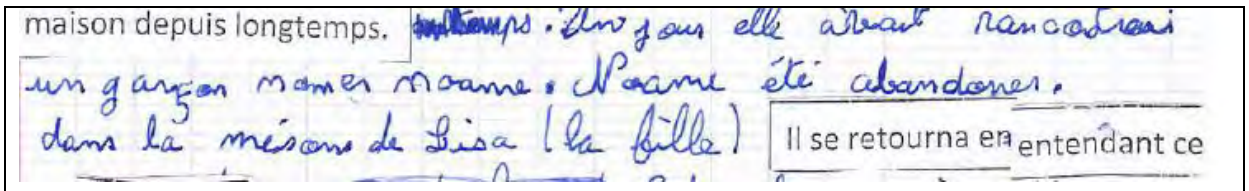

qui habitait dans une joli maison blanche. Elle habitait dans cette maison depuis longtemps. Son frère Eliot entendit grrr. Il se retourna en entendant ce grand bruit. Un

qui habitait dans une jolie maison blanche. *Elle habitait dans cette maison depuis longtemps. S*on frère Eliot entendit « Grrr ». *Il se retourna en entendant ce grand bruit.*

Figure 68 : Texte d'Olivia, CE2 U

La position de sujet syntaxique fait du personnage masculin le thème de l'énoncé et le place donc en position saillante, propice à une reprise du référent dans la phrase suivante dans le cadre d'une progression à thème constant.

Le deuxième type de fonction syntaxique le plus fréquent dans le corpus, quel que soit le niveau de classe, est le complément de verbe : 24% sur l'ensemble du corpus, tel que l'illustre l'exemple (69) avec le complément d'objet direct *un garçon* placé après le verbe *rencontrer* :


maison depuis longtemps. Un jour elle avait rencontré un garçon nomer Noame. Noame été abandoner. dans la mésons de Lisa (la fille) Il se retourna en entendant ce

Un jour, elle avait rencontré un garçon nommé Noam. Noam était abandonné. Dans la maison de Lisa (la fille), *il se retourna en entendant ce grand bruit.*

Figure 69 : Texte de Loulie, CE2 P

La première mention expansée du personnage masculin en position de complément de verbe en fait le rhème de l'énoncé. Dans le cadre d'une progression linéaire, le rhème devient le thème dans l'énoncé suivant : en effet, c'est le cas pour le nom propre *Noam* qui devient le thème de l'énoncé suivant. Par sa position rendue saillante, le lecteur comprend sans difficulté qu'il est le référent du pronom personnel *il* de P2.

9.2. Maintien du référent du pronom personnel *il*

Lorsqu'un interlocuteur fait d'une entité son objet de discours, il se limite rarement à la mentionner une seule fois. Dans son discours, il fera donc appel à des expressions qui reprennent l'entité dont il est question, le *référent*, tout au long du propos et qui sont responsables de la continuité référentielle du texte (Apothéloz, 1995b, Charolles, 2002).

Après avoir été introduit, le référent du pronom personnel *il* peut donc être maintenu plus ou moins longtemps dans le texte. En tant que personnage saillant, on s'attend à ce que le référent du pronom personnel *il* soit mentionné plusieurs fois dans le texte à l'aide de diverses expressions référentielles, constituant ainsi une chaîne de référence (CDR) à la *texture* (Halliday & Hasan, 1976 ; cf. chapitre 1, section 1.3.) importante.

Aussi, en 9.2.1., nous dressons une vue générale du maintien de la référence à travers l'étude du nombre de mentions des CDR et de leur longueur moyenne. En 9.2.2., nous étudions de manière plus précise différents cas de figure, selon qu'il s'agit d'anaphore (1 mention + P1 soit 2 mentions au total), de CDR minimales (2 mentions + P1 soit 3 mentions au total) ou de CDR plus longues (3 mentions ou plus + P1 soit 4 mentions ou plus au total).

9.2.1. Vue d'ensemble des CDR du pronom personnel *il*

De manière générale, l'étude du nombre de mentions (ou maillons) de la CDR du pronom personnel *il* montre que la CDR la plus longue comporte 25 maillons + celui compris dans P2, soit 26 maillons au total, et se trouve au CM1, comme l'illustre le tableau 19. On constate également que les CDR les plus longues (9 à 19 mentions+P2) sont représentées dans chacun des niveaux de classe mais par un seul texte à chaque fois. Les CDR « moyennes » (entre 4 et 8 mentions+P2) représentent entre 8 et 17 textes. Enfin, les CDR les plus courtes (1 à 3 mentions+P2) sont les plus représentées, dans des proportions presque identiques (entre 28 et 31 textes).

Nombre de mentions de la CDR de *il*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
1 mention + P2	14	22%	8	14%	8	12%	30	16%
2 mentions + P2	13	20%	7	13%	11	17%	31	17%
3 mentions + P2	11	17%	4	7%	13	20%	28	15%
4 mentions + P2	5	8%	4	7%	8	12%	17	9%
5 mentions + P2	6	9%	2	4%	3	5%	11	6%
6 mentions + P2	2	3%	6	11%	4	6%	12	6%
7 mentions + P2	1	2%	6	11%	4	6%	11	6%
8 mentions + P2	2	3%	3	5%	3	5%	8	4%
9 mentions + P2	2	3%	0	0%	0	0%	2	1%
10 mentions + P2	0	0%	1	2%	3	5%	4	2%
11 mentions + P2	2	3%	0	0%	1	2%	3	2%
12 mentions + P2	0	0%	1	2%	0	0%	1	1%
13 mentions + P2	0	0%	0	0%	1	2%	1	1%
17 mentions + P2	0	0%	1	2%	1	2%	2	1%
18 mentions + P2	0	0%	0	0%	1	2%	1	1%
19 mentions + P2	0	0%	1	2%	0	0%	1	1%
25 mentions + P2	0	0%	1	2%	0	0%	1	1%
P2 = seule mention	6	9%	11	20%	4	6%	21	11%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 19 : Nombre de mentions dans la CDR incluant le pronom personnel *il*

Dans les textes présentant des CDR plutôt courtes (1 à 3 mentions + P2), on constate qu'au CE2 et au CM1, les cas d'anaphore (1 mention + P2) sont les plus nombreux. En revanche, au CM2, les CDR comportant 3 mentions + P2 sont plus nombreuses que celles n'en comportant qu'une. Il semble donc que plus le niveau scolaire augmente, plus la longueur des CDR augmente. En témoigne la longueur moyenne des CDR incluant le pronom personnel *il*, puisque le nombre moyen de maillons augmente globalement avec le niveau scolaire :

- 4,1 maillons au CE2
- 5,6 maillons au CM1
- 5,3 maillons au CM2

Cependant, le nombre moyen de maillons par CDR est à nuancer car le nombre de CDR longues au CM1 et au CM2 vient influencer la longueur moyenne.

9.2.2. Différentes configurations des CDR

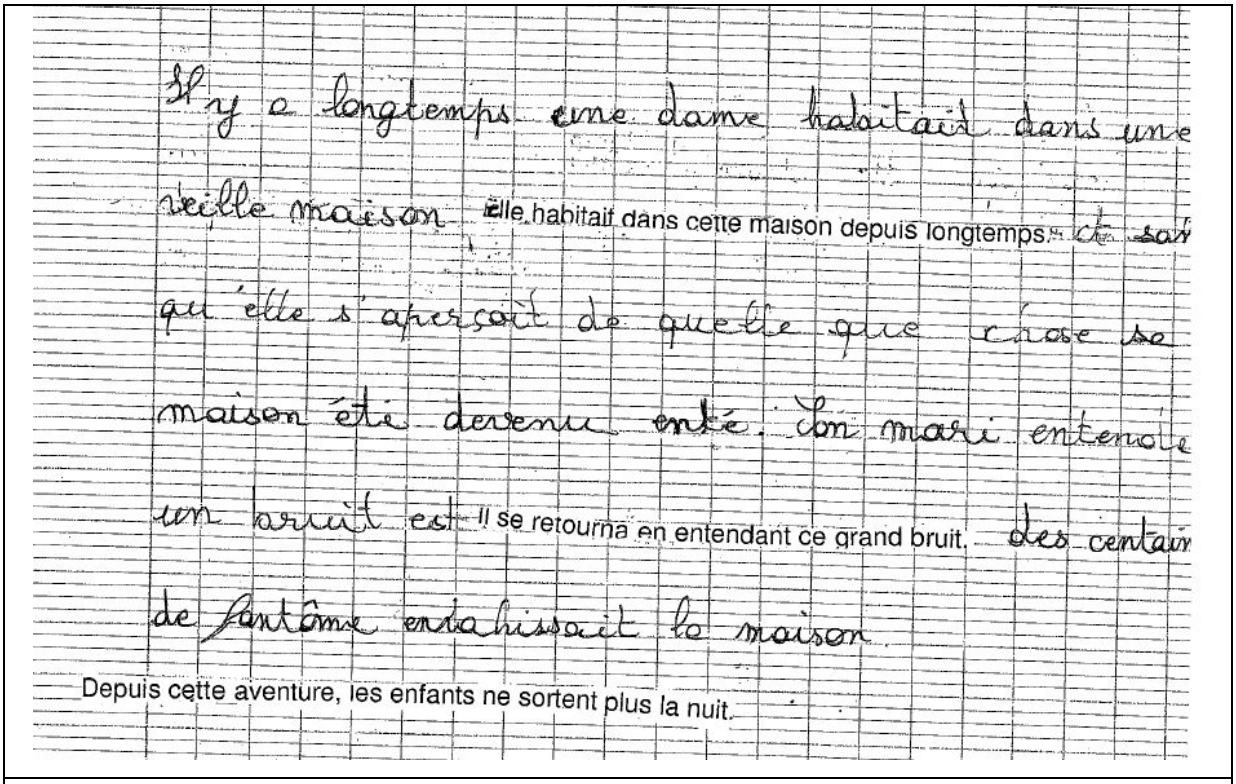
Nous l'avons vu, notre corpus regroupe plusieurs configurations de CDR. Il peut s'agir d'anaphore ou de cataphore (1 mention + P1), de CDR minimales (2 mentions + P1) ou de CDR plus longues (3 mentions ou plus + P1). Chacune de ces configurations fait l'objet d'une étude plus précise afin de savoir si le nombre de mentions d'un référent influe sur les caractéristiques des expressions référentielles utilisées.

9.2.2.1. Étude des cas d'anaphore ou de cataphore : 1 mention + P2

Les textes contenant 1 mention du référent + P2 (soit 2 mentions au total) représentent 16% des textes du corpus (30 textes). Dans cette configuration, on distingue les cas d'anaphore (90%, soit 27 textes) et les cas de cataphore (10%, soit 3 textes).

On parle de cas d'anaphore lorsqu'il y a une mention du référent avant P2 et celle déjà présente dans P2 (Reichler-Béguelin, 1988 ; cf. chapitre 2). Lorsque nous avons étudié l'introduction du référent du pronom personnel *il* (cf. ce chapitre, section 9.1.1.), nous avons vu que la forme de la première mention est le SN indéfini, suivi par le SN possessif (écart de 12 textes entre les deux formes). L'étude de la première mention dans les cas d'anaphore révèle que la forme linguistique privilégiée dans ce cas de figure est le SN possessif, bien que l'écart avec le nombre de SN indéfini soit très mince (1 texte). L'emploi du SN possessif pour introduire le référent quand c'est la seule mention avant P2, peut s'expliquer par l'ordre qu'occupe le pronom personnel *il* de P2 dans la consigne. En effet, le personnage masculin est la troisième occurrence, toutes occurrences confondues, puisqu'il arrive après le pronom personnel *elle* et le SN démonstratif *cette maison*. En terme de saillance, le pronom personnel *il* est la deuxième occurrence parce qu'il occupe la position de sujet syntaxique et de sujet thématique de P2. Il semble donc plus facile d'introduire le référent du pronom personnel *il* en lien avec un référent déjà établi, plutôt que de mobiliser de nouvelles ressources cognitives pour à la fois introduire ce nouveau référent et à la fois gérer son « accrochage » aux éléments du texte déjà construit. En d'autres termes, il semble plus facile d'amener un nouveau personnage grâce à un lien de parenté par exemple avec un autre personnage (d'où l'emploi d'un SN possessif), plutôt que d'introduire un personnage sans lien aucun avec les autres.

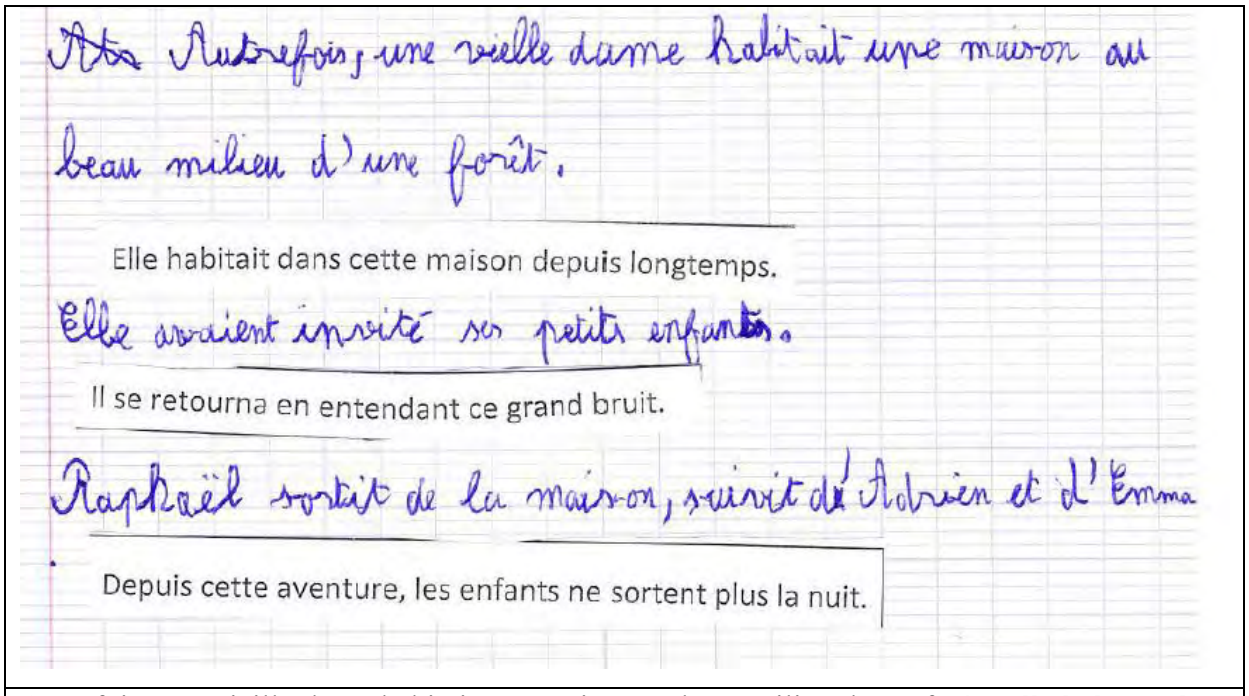
Dans (70) par exemple, la première mention du référent du pronom personnel *il* est le SN possessif *son mari*. Le personnage masculin est donc introduit par le lien familial qui l'unit au personnage féminin, introduit au préalable.


Il y a longtemps une dame habitait dans une
vieille maison. Elle habitait dans cette maison depuis longtemps. Et sans
qu'elle s'aperçoit de quelque chose sa
maison été devenu enté. Son mari entendit
un bruit et Il se retourna en entendant ce grand bruit. des centaine
de fantôme envahissait la maison.
Depuis cette aventure, les enfants ne sortent plus la nuit.

Il y a longtemps une dame habitait dans une vieille maison. *Elle habitait dans cette maison depuis longtemps.* Et sans qu'elle s'aperçoive de quelque chose, sa maison était devenue hantée. Son mari entendit un bruit et *il se retourna en entendant ce grand bruit*. Des centaines de fantômes envahissaient la maison. *Depuis cette aventure les enfants ne sortent plus la nuit.*

Figure 70 : Texte de Noé, CM2 S

On parle de cataphore (cf. chapitre 2, section 2.1.) lorsque le pronom personnel *il* de P2 est la première mention du référent. La source du référent est alors localisée après P2, comme dans (71) avec le nom propre *Raphaël.*


Autrefois, une vieille dame habitait une maison au beau milieu d'une forêt.

Elle habitait dans cette maison depuis longtemps.

Elle avaient invité ses petits enfants.

Il se retourna en entendant ce grand bruit.

Raphaël sortit de la maison, suivit d'Adrien et d'Emma

Depuis cette aventure, les enfants ne sortent plus la nuit.

Autrefois, une vieille dame habitait une maison au beau milieu d'une forêt. *Elle habitait dans cette maison depuis longtemps.* Elle avait invité ses petits-enfants. *Il se retourna en entendant ce grand bruit*. Raphaël sortit de la maison, suivit d'Adrien et d'Emma. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 71 : Texte d'Alban, CE2 P

Les cas de cataphores sont trop peu nombreux (3 textes) et trop variés pour pouvoir tirer des conclusions sur la forme privilégiée de la source. En effet, dans les textes concernés, on trouve un nom propre, un pronom personnel et un SN défini.

En revanche, l'étude des fonctions syntaxiques de la première mention du référent du pronom personnel *il* dans les cas de cataphores comme dans ceux d'anaphores montre que la position syntaxique la plus fréquente est le sujet syntaxique.

9.2.2.2. Étude des CDR minimales : 2 mentions + P2

Nous l'avons vu, un des cas de figure les plus fréquents dans notre corpus lorsqu'il s'agit de maintenir la référence à l'occurrence *il* est le recours à la configuration suivante : 2 mentions + P2 (cf. tableau 19 *supra*), c'est-à-dire 3 mentions au total. Il s'agit là de la configuration minimale dans laquelle on peut parler de CDR (Schnedecker, 1997 ; cf. chapitre

2, section 2.2.). On parle donc de CDR minimale, en opposition avec les CDR contenant plus de 3 maillons que nous aborderons plus tard.

Les CDR minimales témoignent d'une gestion de la référence plus globale. En effet, en mentionnant le référent du pronom personnel *il* plusieurs fois en dehors de P2, l'élève commence à construire son personnage en lui octroyant une place plus importante dans son récit. Les travaux qui étudient le maintien de la référence dans des textes narratifs de scripteurs experts, notamment ceux de Landragin & *al.* (2015 ; cf. chapitre 2, section 2.2.4), montrent qu'une CDR prototypique se réalise le plus souvent sous la forme :

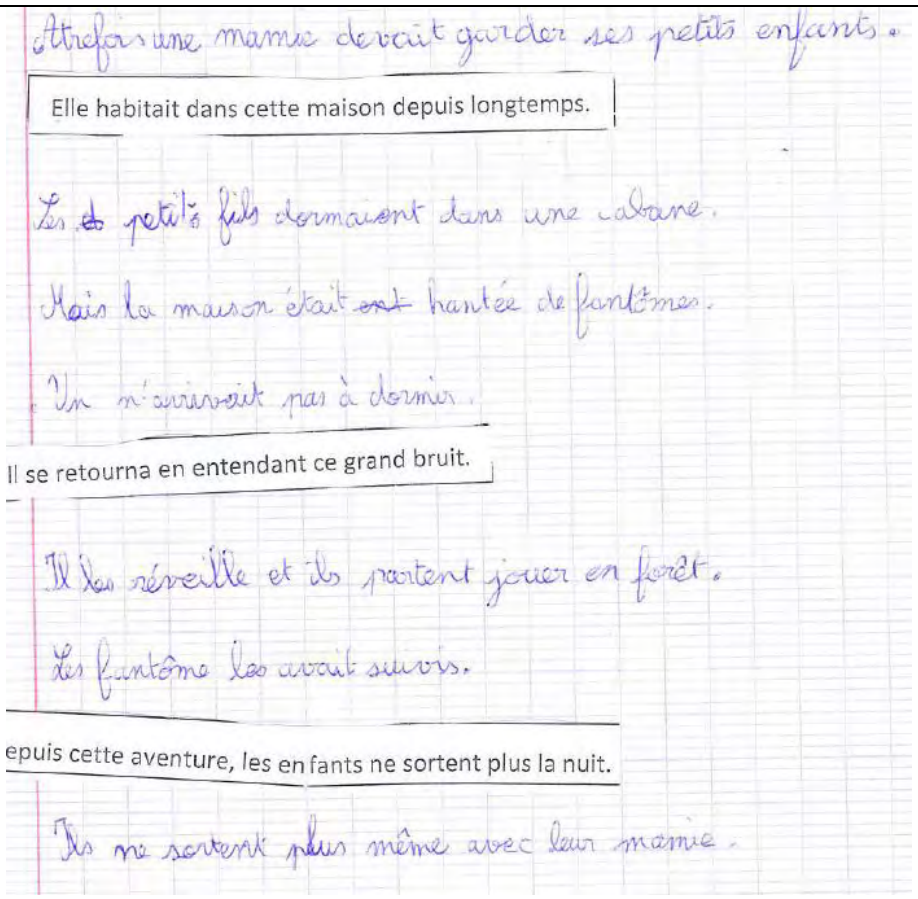
> SN indéfini ou nom propre – description définie – successions de pronoms personnels de troisième personne.

Dans les cas de CDR minimales que nous étudions ici (2 mentions + P2 soit 3 mentions au total), le pronom personnel est présent dans P2. On peut donc s'attendre à ce que la deuxième mention du référent soit majoritairement un SN défini lorsque la mention contenue dans P2 est placée en troisième et dernière position dans la CDR. Lorsque la mention contenue dans P2 est placée en deuxième position dans la CDR, on devrait alors trouver un pronom en troisième mention.

L'étude de la deuxième mention du référent dans les CDR minimales montre que la forme la plus fréquente, quelle que soit la position de P2, est le pronom : 48% des textes contenant une CDR minimale, comme le montrent le tableau 20 et l'exemple (72) :

Formes de M2 dans les CDR minimales de *il*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN défini	1	8%	0	0%	2	18%	3	10%
SN pronominal	8	61%	1	14%	6	55%	15	**48%**
SN possessif	1	8%	2	29%	2	18%	5	16%
SN démonstratif	1	8%	0	0%	0	0%	1	3%
SN indéfini	0	0%	1	14%	0	0%	1	3%
SN nom propre	2	15%	3	**43%**	1	9%	6	19%
Total	13	100%	7	100%	11	100%	31	100%

Tableau 20 : Formes de la deuxième mention du référent du pronom personnel *il* dans les CDR minimales



Autrefois une mamie devait garder ses petits-enfants. *Elle habitait dans cette maison depuis longtemps.* Ses petits-fils dormaient dans une cabane. Mais la maison était hantée de fantômes. Un n'arrivait pas à dormir. *Il se retourna en entendant ce grand bruit.* Il les réveille et ils partent jouer en forêt. Les fantômes les avaient suivis. *Depuis cette aventure, les enfants ne sortent plus la nuit.* Ils ne sortent plus même avec leur mamie.

Figure 72 : Texte de Louise, CE2 P

Malgré le petit nombre de textes concernés, on constate tout de même que si la prédominance du pronom se retrouve dans les niveaux du CE2 et du CM2, le nom propre est majoritaire au CM1 : il est présent dans 43% des cas, quelle que soit la position de P2.

L'étude des fonctions syntaxiques lors de l'introduction du référent (cf. ce chapitre, section 9.1.3.) a montré que la première mention du référent est majoritairement en position de sujet syntaxique. Le personnage masculin de P2 étant une entité saillante, on s'attend à ce qu'il le reste tout au long de la CDR : la deuxième mention du référent devrait également se trouver en position de sujet syntaxique.

Cette hypothèse est confirmée par nos résultats : la deuxième mention des CDR minimales est principalement en position de sujet syntaxique (74% des textes concernés), comme le montre le tableau 21 et l'illustre l'exemple précédent.

Fonctions syntaxiques de M2 dans les CDR minimales de *il*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
sujet	10	77%	4	57%	9	81%	23	74%
complément de verbe	3	23%	3	43%	1	9%	7	23%
complément du nom	0	0%	0	0%	1	9%	1	3%
Total	13	100%	7	100%	11	100%	31	100%

Tableau 21 : Fonctions syntaxiques de la deuxième mention du référent du pronom personnel *il* dans les CDR minimales

9.2.2.3. Étude des CDR plus longues : 3 mentions (ou plus) + P2

Dans les textes comportant des CDR plus longues, c'est-à-dire 3 mentions (ou plus) + celle contenue dans P2, nous rappelons que nous pouvons distinguer (cf. tableau 19) :

- Les CDR comportant 3 mentions + P2 (soit 4 mentions au total), qui représentent 15% des textes du corpus
- Les CDR comportant entre 4 et 8 mentions + P2, qui représentent entre 9% et 4% du corpus
- Les CDR comportant entre 9 et 19 mentions + P2, qui représentent entre 1 et 2 textes chacune et qui figurent globalement dans chaque niveau de classe
- La CDR la plus longue de notre corpus qui comporte 25 mentions + P2 et qui se trouve au CM1

Nous ne pourrions ici faire une revue détaillée de chacune des mentions du référent du pronom personnel *il* comme nous l'avons fait jusqu'à présent. Nous nous contenterons donc d'une vue générale des compositions des CDR comportant 3 maillons (ou plus) + P1.

Quel que soit le nombre de mentions dans la CDR du pronom personnel *il*, le SN pronominal est la forme la plus fréquemment employée pour maintenir la référence. Ce

résultat est en accord avec les différents travaux qui s'intéressent au maintien de la référence chez les enfants. En effet, Karmiloff-Smith (1980, 1981 ; cf. chapitre 3, section 3.1.2.2.) notamment, montre que les pronoms sont les formes privilégiées de la reprise, d'autant plus lorsque le sujet grammatical est thématisé. Le pronom est la forme « non marquée » de la reprise (Kleiber, 1994 ; cf. chapitre 2, section 2.2.4.), donc la plus facilement et longtemps utilisable, jusqu'à la nécessité d'une redénomination. On constate toutefois, dans les chaînes les plus longues, quelques formes différentes qui alternent avec le pronom, telles que des noms propres, des SN définis ou des formes atténuées de la référence comme les pronoms zéro.

En ce qui concerne les expansions des mentions du référent, on trouve des formes expansées à l'aide d'un modifieur dans toutes les CDR comportant jusqu'à 11 mentions + P2, même s'il s'agit d'un ou deux textes dans chaque type de CDR. Les différents types d'expansion (syntagme adjectival, SN prépositionnel, SN nom propre) sont tous représentés, à l'exception des propositions relatives.

Enfin, l'étude des fonctions syntaxiques des différentes mentions montre qu'elles sont le plus fréquemment en position de sujet syntaxique, à l'image de la position syntaxique occupée par le pronom personnel *il* dans P2 et qui fait du personnage masculin le sujet thématique. Les élèves vont donc avoir tendance à le thématiser également dans les mentions antérieures. C'est la stratégie du sujet thématique (Karmiloff-Smith, 1980, 1981 ; cf. chapitre 3, section 3.1.2.3.).

Nous retrouvons ces éléments dans le texte d'Alix (CM1 T1), que nous reproduisons dans son intégralité et qui comporte la plus longue CDR incluant le pronom personnel *il* dans notre corpus, puisqu'elle comporte 25 mentions du référent + P2.

À côté d'un zoo il y avait une maison, dans la maison il y avait une dame,

Elle habitait dans cette maison depuis longtemps.

Dan le zoo, il y avait un T. rex, mais il y avait des grandes clôtur électrique pour l'empêcher de sortir. La dame n'avait jama visiter le zoo, elle appelle ses enfants pour aller au zoo avec eux. Jeudi, ils vont au zoo, ils visitent le zoo tout entier, et il voient le T. rex. Heureusement, ils visitaient le zoo en voiture ils ont eux peur du dinosaure et ils sont vite partit, en démarrant ils ont fait du bruit, le T. rex a entendu le bruit et

Il se retourna en entendant ce grand bru

Le dinosaure a vu la voiture et il a essayé de l'attraper, mais il n'

pas réussi. Le T. rex était très fâché de ne pas l'avoir attraper. Il était tellement en colère qu'il déchira la clôture, il s'est fait un peu électrifier, mais il s'en est vite remis, la voiture était déjà loin. Le dinosaure avait de grandes jambes et il a vite ratrapé le véhicule, il se plaça devant la voiture, la dame a vis freiné, elle regarda le T. rex et elle s'est enfuie. Le dinosaure l'a attrapé et il l'a mangé, les enfants avait peur et ils se faisaient discret dans la voiture. Le T. rex soulevra l'auto et il l'envoya très loin. Les gardiens du zoo sont alerté par le bruit de la voiture, ils appellent une ambulance et les policie les deux enfants sont transporter à l'hôpital. Les policiers et les gardiens qui sont resté au zoo vont voir l'endos du T. rex, ils voient que la clôture est déchiré, les policiers vont chercher leurs voitures et ils partent à la recherche du dinosaure, le T. rex attrape les policiers et ils les mange

Depuis cette aventure, les enfants ne sortent plus la nuit.

A côté d'un zoo il y avait une maison, dans la maison il y avait une dame, *elle habitait dans cette maison depuis longtemps.* Dans le zoo, il y avait un T-Rex, mais il y avait des grandes clôtures électriques pour l'empêcher de sortir. La dame n'avait jamais visité le zoo, elle appelle ses enfants pour aller au zoo avec eux. Jeudi, ils vont au zoo, ils visitent le zoo tout entier, et ils voient le T-Rex. Heureusement, ils visitaient le zoo en voiture. Ils ont eu peur du dinosaure et ils sont vite partis. En démarrant, ils ont fait du bruit, le T-Rex a entendu le bruit et *il se retourna en entendant ce grand bruit*. Le dinosaure a vu la voiture et il a essayé de l'attraper, mais il n'a pas réussi. Le T-Rex était très fâché de ne pas l'avoir attrapée. Il était tellement en colère qu'il déchira la clôture, il s'est fait un peu électrifier, mais il s'en est vite remis. La voiture était déjà loin. Le dinosaure avait de grandes jambes et il a vite rattrapé le véhicule. Il se plaça devant la voiture, la dame a vite freiné, elle regarda le T-Rex et elle s'est enfuie. Le dinosaure l'a attrapée et il l'a mangée. Les enfants avaient peur et ils se faisaient discrets dans la voiture. Le T-Rex souleva l'auto et il l'envoya très loin. Les gardiens du zoo sont alertés par le bruit de la voiture, ils appellent une ambulance et les policiers. Les deux enfants sont transportés à l'hôpital. Les policiers et les gardiens qui sont restés au zoo vont voir l'enclos du T-Rex, ils voient que la clôture est déchirée. Les policiers vont chercher leurs voitures et ils partent à la recherche du dinosaure, le T-Rex attrape les policiers et il les mange. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 73 : Texte d'Alix, CM1 T1

Dans ce texte, le référent du pronom personnel *il* est mentionné 5 fois avant P2. La forme de la première mention, le SN indéfini *un T-Rex* (ligne 3), est conforme à ce qui est attendu. En effet, étant donné que c'est la première fois que l'entité apparaît et que son accessibilité est faible pour le lecteur, l'emploi d'un SN indéfini est justifié. La deuxième mention du référent sous la forme du pronom *l'* dans *pour l'empêcher de sortir* (ligne 4) est là aussi justifiée puisque l'accessibilité du référent est devenue forte, le pronom est donc la forme privilégiée de la reprise. Quelques lignes plus loin, alors que le personnage du dinosaure semble être passé à l'arrière-plan, il est réinstancié à l'aide du SN défini *le T-Rex* (ligne 7) puis repris dans la ligne suivante par l'hyperonyme *dinosaure*, afin d'éviter les répétitions. On constate que ces formes sont en position de complément de verbe. Cela paraît pertinent puisque le focus d'attention est centré sur le personnage féminin étant donné que c'est le premier à être introduit dans le texte et que c'est celui dont on parle principalement. C'est donc le thème grammatical et le sujet syntaxique.

Le personnage du dinosaure devient le sujet thématique à partir de la 5e mention, juste avant P2. Le SN défini *le T-Rex* devient alors le sujet syntaxique de la phrase (ligne 9) et le reste quasiment jusqu'à la fin du texte. Seules les formes de la reprise varient en une

alternance de pronoms et de SN définis *le T-Rex* ou *le dinosaure*. La chaîne de référence du personnage masculin croise à plusieurs reprises celle du personnage féminin et celle des enfants, en des relations d'entrecroisement (Schnedecker, 1997 ; cf. chapitre 2, section 2.2.2.). Ces croisements se traduisent par un changement de position syntaxique des référents des personnages. Par exemple, lorsque le personnage féminin revient dans le focus d'attention, les mentions du référent au pronom personnel *elle* repassent en position de sujet syntaxique et le SN défini *le dinosaure* se retrouve en complément de verbe (mention17) : *la dame a vite freiné, elle regarda le T-Rex* (lignes 15-16).

9.3. Bilan de l'étude du référent du pronom personnel *il*

De manière générale, l'étude des formes de la première mention du référent du pronom personnel *il* montre que :

- la forme privilégiée pour introduire le référent est le SN indéfini, avec une fréquence d'emploi croissante en fonction du niveau scolaire. Toutefois, les élèves les plus jeunes emploient également le SN possessif en première mention, dans les mêmes proportions.
- la première mention du référent apparaît majoritairement en position de sujet syntaxique, plus rarement en position de complément de verbe, quel que soit le niveau de classe.
- les formes lexicales des SN employés en première mention sont principalement des noms relationnels en lien avec la cellule familiale : *mari, père, fils, petit-frère, grand-père,* quel que soit le niveau de classe. Les noms génériques (*garçon, homme*) sont moins nombreux et les noms fonctionnels (*chasseur, marchand, voisin*) minoritaires.
- lorsque la forme de la première mention est expansée, elle l'est principalement à l'aide d'une seule expansion, environ dans les mêmes proportions quel que soit l'âge des élèves.
- les types d'expansion utilisés varient en fonction du niveau scolaire : les propositions relatives sont plus nombreuses lors que les élèves sont plus jeunes, les syntagmes adjectivaux sont plus nombreux lorsque les élèves sont plus âgés.

L'étude du maintien de la référence au pronom personnel *il* montre plusieurs manières de « gérer » la référence. La première est une gestion plutôt locale du référent : les textes présentent deux mentions du référent, c'est-à-dire 1 mention + celle de P2. Dans cette configuration, la mention employée par les élèves (ou source) est localisée soit avant P2 (anaphore), soit après P2 (cataphore). Dans les cas d'anaphore, il y a presque autant de SN indéfinis employés en première mention que de SN possessifs. En outre, quelle que soit la localisation de la source, la fonction syntaxique la plus fréquemment utilisée est le sujet.

La deuxième manière de « gérer » la référence se traduit par une gestion plus globale du référent : les élèves le mentionnent plusieurs fois dans leur texte, ce qui constitue des CDR plus ou moins longues. Lorsque la CDR déployée pour le pronom personnel *il* est minimale (2 mentions + P1, soit 3 mentions au total), ou bien plus longue (3 mentions ou plus + P1), on constate une prédominance de SN pronominaux comme formes de reprise du référent, en position de sujet syntaxique, excepté lorsque la chaîne de référence du personnage masculin entretient des relations d'entrecroisement avec des chaînes référant à d'autres personnages. Les formes des mentions peuvent également varier (SN définis, noms propres notamment) en fonction des nécessités de réinstanciation du personnage, que ce soit pour des raisons de saillance ou de changement de personnage dans le focus d'attention en raison de l'entrecroisement de plusieurs chaînes de référence référant à plusieurs personnages.

CHAPITRE 10 : Introduction et maintien du référent du SN démonstratif *cette maison*

Dans la tâche-problème que doivent réaliser les élèves de notre étude, la première phrase à insérer (P1) est la suivante :

Elle habitait dans cette maison depuis longtemps.

Dans cette phrase, l'emploi du SN démonstratif *cette maison* demande un traitement particulier de la part des élèves. En effet, les SN démonstratifs sont des expressions déictiques qui nécessitent le recours à la situation d'énonciation pour être interprétés (Charolles, 2002 ; cf. chapitre 1, section 1.4.2.1.). Le démonstratif dirige l'attention sur une entité identifiable car immédiatement perceptible dans la situation d'énonciation (geste ostensif à l'oral par exemple). Les SN démonstratifs possèdent également une valeur anaphorique, comme c'est le cas dans P1. Pour être interprétés, les SN démonstratifs anaphoriques nécessitent le recours à une entité identifiable car venant d'être évoquée dans le discours au moyen d'expressions linguistiques. Il y a donc la présence d'une source avec laquelle le SN démonstratif peut être mis en relation car « l'interprétation anaphorique d'un SN démonstratif […] requiert l'occurrence effective dans le contexte gauche d'un terme avec lequel [il peut] être mis en relation […] » (Marandin, 1986 : 77). Dans les textes d'apprentis scripteurs, les deux valeurs des SN démonstratifs (déictique et anaphorique) peuvent être la source de tensions car elles ne sont pas encore maîtrisées.

Aussi, en 10.1., nous allons étudier la manière dont les élèves introduisent le référent du SN démonstratif *cette maison*, en analysant successivement les formes de la première mention du référent, leurs éventuelles expansions et leur fonction syntaxique. Après avoir introduit le référent, les élèves peuvent le maintenir tout au long du discours, ou tout du moins dans une partie du discours. En 10.2., nous allons, de ce fait, étudier la manière dont les élèves maintiennent la référence tout au long du texte.

10.1. Introduction du référent du SN démonstratif *cette maison*

De manière générale, dans notre corpus, nous observons trois localisations possibles de l'expression référentielle qui introduit le référent (ou *source*), comme le montre le tableau 22 :

- La source est introduite avant la phrase imposée et la référence est maintenue après
- La source n'est pas introduite avant la phrase imposée mais elle est localisée après
- La source est absente : la seule mention du référent est celle de la phrase imposée

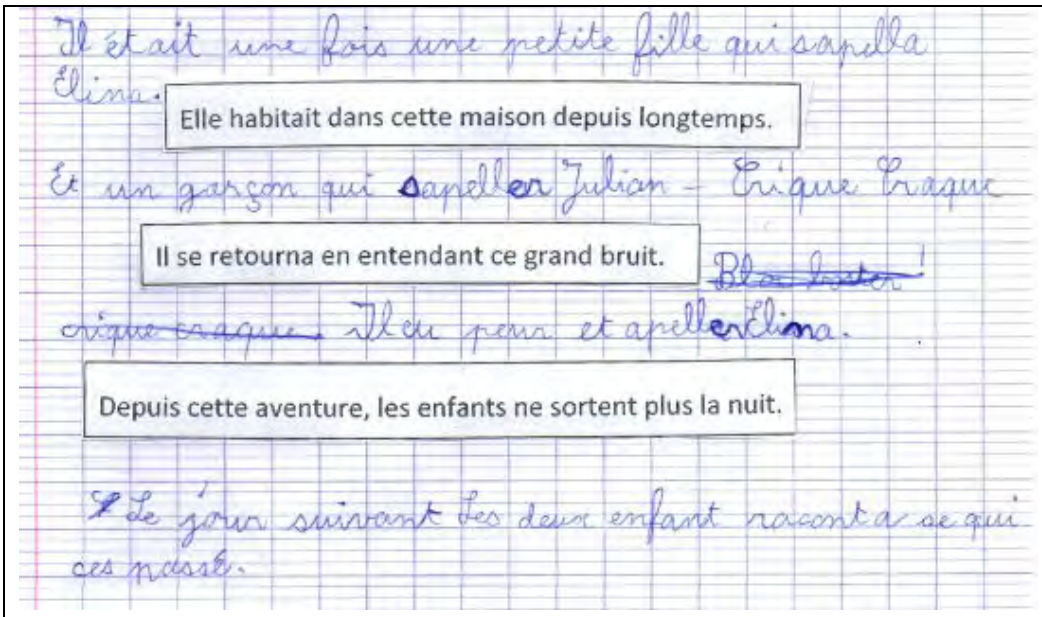
Localisation de la source du SN *cette maison*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
avant P1 et maintien après	38	59%	31	55%	45	69%	114	**62%**
après P1	9	14%	7	13%	7	11%	23	**12%**
source absente	17	27%	18	32%	13	20%	48	**26%**
Total	64	100%	56	100%	65	100%	185	100%

Tableau 22 : Localisation de la source du SN *cette maison*

Sur l'ensemble du corpus, dans 62% des textes le référent est introduit avant P1 et maintenu par la suite dans le discours. Globalement, les textes présentant cette configuration sont plus nombreux au CM2 que dans les autres niveaux de classe : 59% au CE2, 55% au CM1, 69% au CM2.

De manière plus marginale, dans 12% des textes du corpus, la source du SN démonstratif *cette maison* est localisée après P1. Il s'agit de cataphores (cf. chapitre 2, section 2.1.1.1.) : le SN démonstratif *cette maison* de P1 est la première mention du référent, qui est repris par la suite sous diverses formes. Cette manière d'introduire le référent est plus fréquente lorsque les élèves sont plus jeunes : 14% au CE2, 13% au CM1 et 11% au CM2.

Enfin, sur l'ensemble du corpus, dans 26% des textes la source du SN démonstratif *cette maison* est absente, comme dans (74) :



Il était une fois une petite fille qui s'appelait Elina. *Elle habitait dans cette maison depuis longtemps.* Et un garçon qui s'appelait Julian. « Crique craque ». *Il se retourna en entendant ce grand bruit.* Il eut peur et appela Elina. *Depuis cette aventure, les enfants ne sortent plus la nuit.* Le jour suivant, les deux enfants racontèrent ce qu'il s'est passé.

Figure 74 : Texte de Julian, CE2 L

On pourrait s'attendre à ce que la fréquence de cette absence décroisse avec l'âge dans la mesure où plus le niveau de classe augmente, plus les élèves savent à la fois introduire et maintenir un référent dans un discours (cf. chapitre 3). Or on remarque au CM1 une légère différence par rapport aux autres niveaux : dans 32% des textes au CM1 le référent du SN démonstratif *cette maison* n'est pas introduit, contre 27% au CE2 et 20% au CM2.

Après avoir dressé une vue d'ensemble des différentes manières de gérer la référence, nous allons étudier les formes de la première mention du référent du SN démonstratif *cette maison*.

10.1.1. Etude des formes de la première mention du référent du SN démonstratif *cette maison*

Le SN démonstratif *cette maison* implique un haut degré d'accessibilité du référent. Pour que le lecteur comprenne à quelle entité réfère le SN démonstratif, l'introduction d'un

référent préalablement à P1 s'impose. Le référent ainsi introduit constitue la première mention (M1) du SN démonstratif *cette maison.* Le tableau 23 reproduit les différents types de formes sous lesquelles se réalise la première mention.

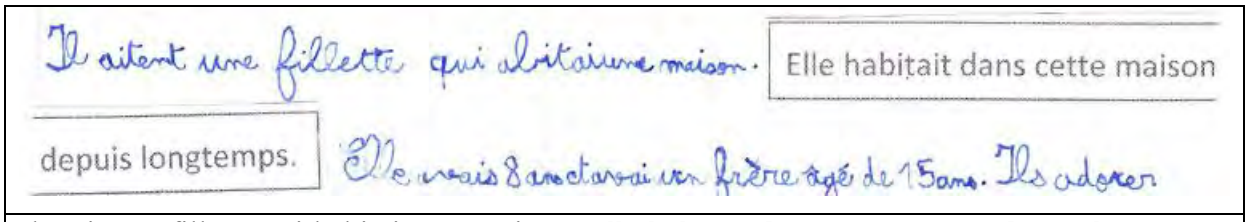
Formes de M1 pour *cette maison*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN indéfini	26	41%	19	34%	31	48%	76	**41%**
SN défini	7	11%	10	18%	4	6%	21	**11%**
SN possessif	1	2%	4	7%	4	6%	9	5%
SN démonstratif	3	5%	1	2%	4	6%	8	4%
SN pronominal	1	2%	0	0%	0	0%	1	1%
autre	8	13%	4	7%	9	14%	21	**11%**
P1 = seule mention	18	28%	18	32%	13	20%	49	26%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 23 : Formes de la première mention du référent du SN démonstratif *cette maison*

Les différentes formes d'introduction du référent du SN démonstratif cette maison

Les travaux qui se sont intéressés à l'introduction des référents dans un discours chez les enfants ou les adultes aboutissent à la conclusion que l'introduction d'un référent par un SN indéfini se réalise progressivement et tardivement chez les enfants. En outre, si les enfants emploient majoritairement un SN indéfini pour introduire le référent, ils ne le font pas de manière systématique. Les recherches de Warden (1976 ; 1981 ; cf. chapitre 3, section 3.1.2.1.) par exemple, montrent qu'il faut attendre l'âge de 9-10 ans (CM1) pour que les enfants introduisent exclusivement les référents par un SN indéfini. Avant 9-10 ans, on constate des introductions au moyen de défini ou d'indéfini.

Nous retrouvons ces résultats dans nos propres travaux, dans une moindre mesure. En effet, la première mention du référent est principalement introduite sous la forme d'un SN indéfini : 41 % dans l'ensemble du corpus, comme l'illustre l'exemple (75) où *cette maison* de P1 a pour source le SN indéfini *une maison* dans la phrase précédente :

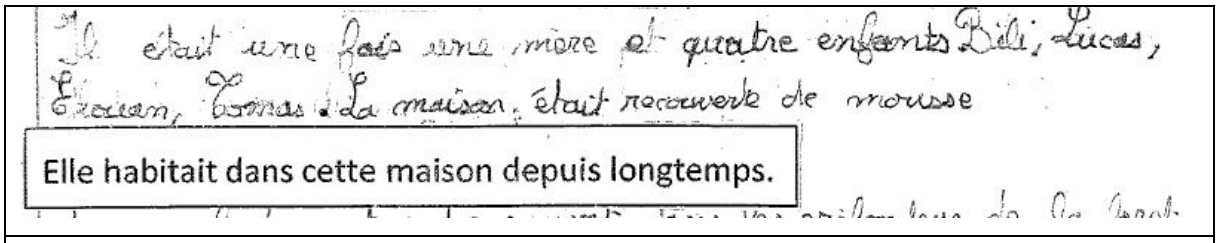


Il était une fillette qui habitait une maison. *Elle habitait dans cette maison depuis longtemps.* Elle avait 8 ans et avait un frère âgé de 15 ans.

Figure 75 : Texte d'Albane, CE2 T

Les travaux précédemment cités montrent une maîtrise progressive de l'introduction des référents par un SN indéfini. De manière globale, dans notre corpus, les élèves les plus âgés sont effectivement ceux pour lesquels la fréquence d'emploi d'un SN indéfini est la plus grande : 48% au CM2 contre 41% au CE2.

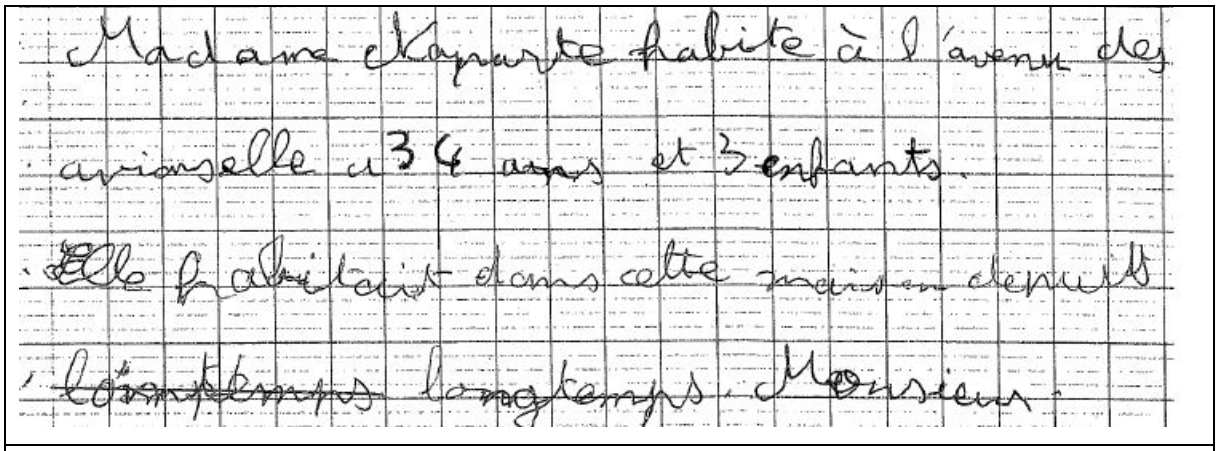
La deuxième forme la plus fréquente pour introduire le référent du SN démonstratif *cette maison* est le SN défini : 11% dans l'ensemble du corpus, comme dans (76).



Il était une fois une mère et quatre enfants Bili, Lucas, Elouan, Tomas. La maison était recouverte de mousse.

Figure 76 : Texte de Camille, CM1 T1

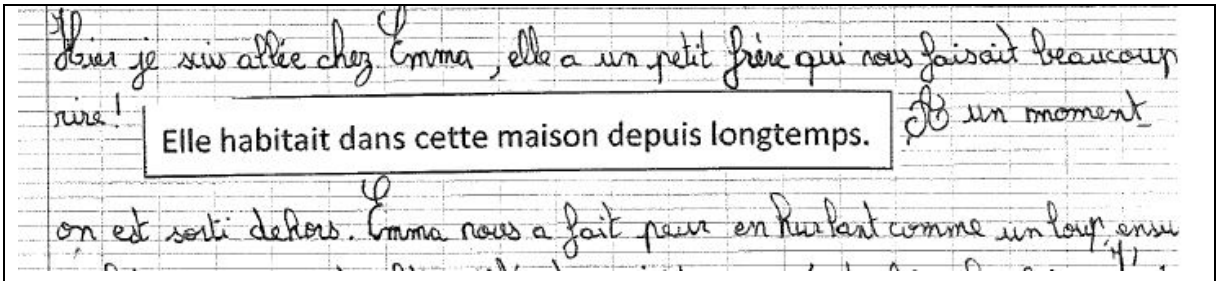
Dans les différents types de forme employées pour la première mention du référent du SN démonstratif *cette maison*, la catégorie *autre* représente 11% des textes de notre corpus. Il s'agit de textes dans lesquels le lecteur doit procéder à un calcul interprétatif pour comprendre la référence au SN démonstratif. Par exemple dans (77), l'emploi du verbe *habiter* fait appel aux connaissances partagées par le lecteur : dans le monde dans lequel nous vivons, il est naturel de penser que si une personne *habite* quelque part, elle habite dans une maison. En outre, le nom *avenue* peut confirmer qu'il s'agit bien d'une maison puisqu'on imagine sans peine qu'une avenue est bordée de maisons ou de grands bâtiments.



Madame Naparte habite à l'avenue des avions, elle a 34 ans et 3 enfants. *Elle habitait dans cette maison depuis longtemps.*

Figure 77 : Texte de Nicolas, CM2 T

Le calcul interprétatif peut également s'appuyer sur l'indication de localisation spatiale contenue dans le SN prépositionnel *chez machin*, comme dans (78) où l'on comprend que *chez Emma* signifie *dans la maison d'Emma* :



Hier je suis allée chez Emma, elle a un petit frère qui nous faisait beaucoup rire ! *Elle habitait dans cette maison depuis longtemps.* À un moment, on est sorti dehors.

Figure 78 : Texte de Pauline, CM1 T1

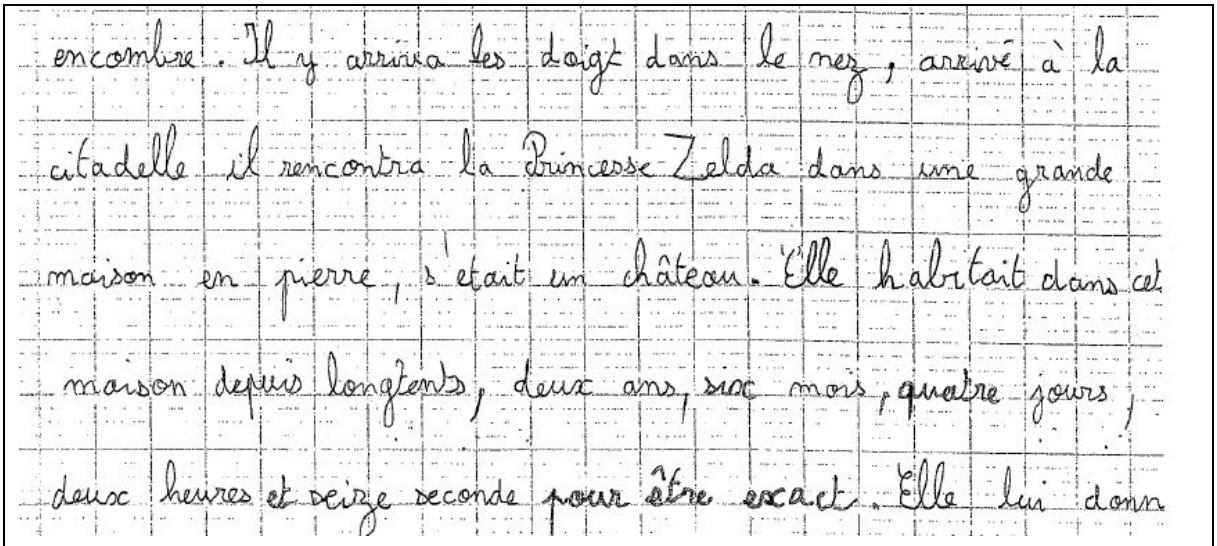
Enfin, quelques formes de la première mention du SN démonstratif *cette maison* apparaissent de manière minoritaire dans le corpus. Il s'agit de SN possessifs (9 textes) tels que *sa maison*, de SN démonstratifs (8 textes) comme *cette maison* ou bien de SN pronominaux (1 texte) tel que le pronom personnel *elle*.

Étude des formes lexicales de M1 pour cette maison

L'étude des formes lexicales de la première mention du référent montre que la forme *maison* est la plus fréquente dans le corpus (cf. exemples cités *supra*). Les élèves gardent le terme générique déjà employé dans P1, quel que soit le niveau de classe.

Cependant, dans une minorité de textes du corpus (de l'ordre d'un ou deux pour chacune des formes), on trouve des termes plus spécifiques faisant appel au type de construction tels que *manoir, château,* ou même *maison de retraite*. Ce type de forme lexicale se trouve au CM1 et au CM2. L'emploi de ces termes peut parfois poser problème pour la cohérence textuelle. En effet, le sémantisme du nom tête *maison* du SN démonstratif *cette maison* est directement relié au verbe *habiter* qui le précède. La cooccurrence du verbe *habiter* et du SN *maison* dans P1 peut appeler deux types de sens lexicaux. Soit on comprend P1 comme évoquant le foyer, dans le sens de *home* en anglais *(sweet home)*, soit on la comprend comme évoquant le type de bâtiment, dans le sens de *house*.

En outre, dans certains textes, on trouve l'emploi de différentes unités lexicales pour référer au SN démonstratif *cette maison*. Les diverses unités utilisées entretiennent entre elles des liens d'hyponymie/hyperonymie. Mais lorsque le SN employé est mal choisi, il peut devenir un obstacle à la cohérence textuelle, comme dans (79) :


encombre. Il y arriva les doigt dans le nez, arrivé à la citadelle il rencontra la Princesse Zelda dans une grande maison en pierre, c'était un château. Elle habitait dans cet maison depuis longtemps, deux ans, six mois, quatre jours, deux heures et seize seconde pour être exact. Elle lui donn

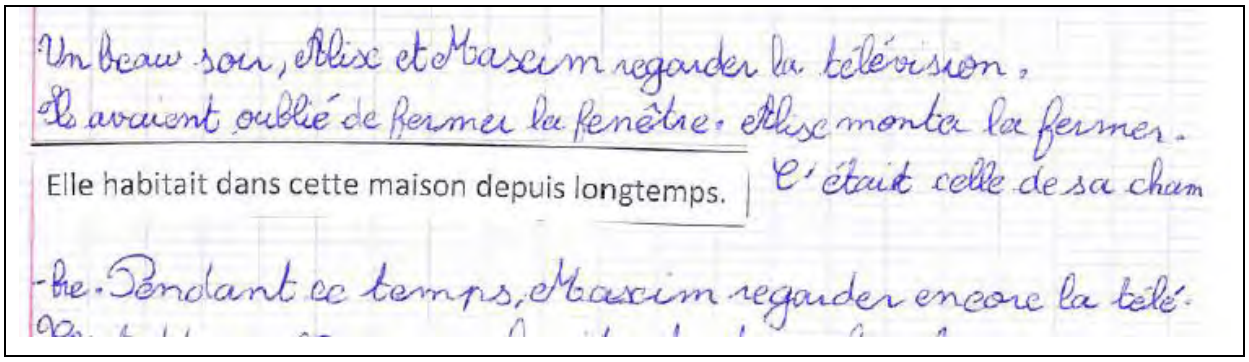
[encombre]. Il y arriva les doigts dans le nez. Arrivé à la citadelle, il rencontra la princesse Zelda dans une grande maison en pierre, c'était un château. *Elle habitait dans cette maison depuis longtemps*, deux ans, six mois, quatre jours, deux heures et seize secondes pour être exact. Elle lui donna…

Figure 79 : Texte d'Antoine, CM2 T

La première mention du référent du SN démonstratif *cette maison* est le SN indéfini *une maison en pierre*. L'unité lexicale *maison* est générique, c'est un hyperonyme. La deuxième mention du référent se trouve dans la même phrase, sous la forme de l'hyponyme *château*. Puis dans P1, qui est la phrase suivante, on a de nouveau l'hyperonyme *maison*. Le passage d'un terme générique à un terme spécifique puis de nouveau au terme générique peut être une gêne pour la cohérence du texte car comme le montre Charolles (1978), il vaut mieux débuter par un terme spécifique puis après seulement, employer un terme générique, l'inverse paraissant moins naturel.

Dans les formes lexicales employées pour référer au SN démonstratif *cette* maison, nous remarquons deux cas particuliers. Ce sont tous deux des textes de CE2 dans lesquels on peut penser que la première mention du référent du SN démonstratif *cette maison* est un nom référent à une partie de la maison, comme dans les anaphores associatives. Pour rappel, l'anaphore associative ne « reprend » aucun élément explicitement mentionné ailleurs dans l'énoncé ou le contexte. L'anaphore dans ce cas, ne s'accompagne pas de coréférence mais se caractérise par le fait qu'elle « rappelle » une information accessible, présentée comme connue, mais non directement fournie par le contexte linguistique et sans indication de rapport aux autres référents ou aux autres informations explicitement formulées (Reichler-Béguelin, 1989 ; Kleiber, 1994 ; cf. chapitre 2, section 2.1.2.1.). Pour donner un sens référentiel aux éléments de ce type d'anaphore, il faut recourir au savoir partagé (Berrendonner, 1983, 1986 ; Reichler-Béguelin, 1989) d'une communauté donnée qui autorise les raisonnements implicites, savoir partagé « relatif aux rapports de type méréologique et métonymique qui existent entre les réalités dénotées, dont certaines peuvent ainsi devenir objets de discours sans avoir été introduites au préalable » (Reichler-Béguelin, 1988a : 19).

Dans (80) par exemple, on peut s'appuyer sur plusieurs indices pour réaliser le calcul interprétatif nécessaire.


Un beau soir, Alix et Maxim regarder la télévision.
Ils avaient oublié de fermer la fenêtre. Alix monta la fermer.
Elle habitait dans cette maison depuis longtemps.
C'était celle de sa cham
-bre. Pendant ce temps, Maxim regarder encore la télé.

Un beau soir, Alix et Maxime regardaient la télévision. Ils avaient oublié de fermer la fenêtre. Alix monta la fermer. *Elle habitait dans cette maison depuis longtemps.* C'était celle de sa chambre. Pendant ce temps, Maxime regardait encore la télé.

Figure 80 : Texte de Noémie, CE2 P

Le SN défini *la fenêtre* est en lien direct avec le SN démonstratif *cette maison* puisque dans le monde dans lequel nous vivons, une maison est composée de 4 murs, d'un toit, de portes et de fenêtres. En outre, le verbe *monter* suggère la présence d'un escalier et donc d'un étage, élément également caractéristique d'une maison (on aurait plus de mal à admettre la présence d'un étage dans un appartement par exemple, bien qu'il existe des duplex). Si un doute subsistait, on pourrait également s'appuyer sur la présence de la *télévision* pour comprendre que les personnages sont à l'intérieur d'un bâtiment. Pour autant, on peut se demander si tous ces éléments sur lesquels s'appuie le calcul interprétatif du lecteur suffisent pour dire que le référent est bien introduit. Il nous semble que la réponse à cette interrogation dépend du degré de coopération du lecteur et donc, d'une part de subjectivité en lien avec l'interprétabilité des textes (Charolles, 1988a ; cf. chapitre 1, section 1.1.). La posture du lecteur, à savoir s'il s'agit d'un enseignant, d'un linguiste ou d'un parent d'élève par exemple, peut également influer sur son degré de coopération car comme le montre Masseron (1989) au sujet des enseignants : « le professeur ne lit pas, il corrige ».

Ainsi, l'étude des formes linguistiques employées pour la première mention du référent du SN démonstratif *cette maison* montre que la forme la plus employée est le SN indéfini mais que des textes présentent également des SN définis et des éléments nécessitant un calcul interprétatif. La forme lexicale la plus utilisée est l'unité *maison* toutefois, quelques textes emploient des noms spécifiques tels que *château, manoir, cabane*.

10.1.2. Etude des expansions de la première mention du référent du SN démonstratif *cette maison*

Lors de l'introduction du référent du SN démonstratif *cette maison* présent dans P1, on trouve dans les textes des SN nus et des SN expansés, présentant un ou plusieurs modifieurs qui sont autant d'ajouts d'informations à propos de l'entité ainsi dénotée (Charolles, 2002).

L'étude des formes de la première mention du référent du SN démonstratif *cette maison* montre que si, dans leur majorité, les textes du corpus ne présentent pas d'expansion, ceux qui en présentent une seule (25%) sont plus nombreux que ceux qui en présentent deux (7 textes) ou trois (1 seul texte), comme le montre le tableau 24. En outre, il semble que plus les élèves sont jeunes, moins ils ont tendance à employer d'expansion.

Nombre d'expansions de M1 pour *cette maison*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
1 expansion	13	20%	13	23%	21	32%	47	**25%**
2 expansions	2	3%	2	4%	3	5%	7	**4%**
3 expansions	0	0%	1	2%	0	0%	1	**1%**
pas d'expansion	49	77%	40	71%	41	63%	130	70%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 24 : Nombre d'expansions de la première mention du référent du SN démonstratif *cette maison*

Étude des types de première expansion dans la première mention du référent du SN démonstratif cette maison

Les différents types d'expansions utilisés peuvent être de différentes sortes : syntagme adjectival, proposition relative, syntagme prépositionnel ou nom propre, comme on le voit dans le tableau 25.

Type de première expansion de M1 pour *cette maison*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
syntagme adjectival	11	17%	12	21%	19	29%	42	**23%**
syntagme prépositionnel	3	5%	3	5%	5	8%	11	**6%**
proposition relative	0	0%	1	2%	0	0%	**1**	1%
SN nom Propre	1	1%	0	0%	0	0%	**1**	1%
non pertinent	49	77%	40	71%	41	63%	130	70%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 25 : Type de première expansion dans les formes de la première mention du référent du SN démonstratif *cette maison*

L'étude des types d'expansions de la première mention du référent du SN démonstratif *cette maison* montre que lorsque le SN introduisant la première mention est expansé, il l'est le plus fréquemment à l'aide d'un syntagme adjectival : 23% sur l'ensemble du corpus, et la fréquence d'emploi d'un SN adjectival augmente en fonction du niveau de classe. Les formes lexicales les plus rencontrées sont : *petite, grande* ou *vieille,* comme on le voit dans (81) :

C'étais une veille dame qui des enfant dans un
veille maison. Elle habitait cette maison depuis
longtemps. Est cette maison etait très très veille.
Elle adorait garder des enfants c'était sa pas

C'était une vieille dame qui *(élevait ?)* des enfants dans une vieille maison. *Elle habitait dans cette maison depuis longtemps.* Et cette maison était très très vieille. Elle adorait garder des enfants, c'était sa passion.

Figure 81 : Texte de Carlo, CM2 T

Parmi les autres types d'expansion, on trouve également :

- des syntagmes prépositionnels : 6% du corpus. Il s'agit de formes telles que *une maison à côté du village* (Aurore, CE2 P) ou *une maison en bambou* (Naélia, CM1 T1)
- une proposition relative : *une maison qui était au beau milieu d'un champ* (Corentin, CM1 T1)
- un SN nom propre : *une maison appelée FKG* (Eliott 2, CE2 T)

Comme on le voit à travers les exemples en italique, les informations sémantiques apportées par les modifieurs utilisés sont variées et peuvent porter sur l'état de la maison, sa localisation, son nom.

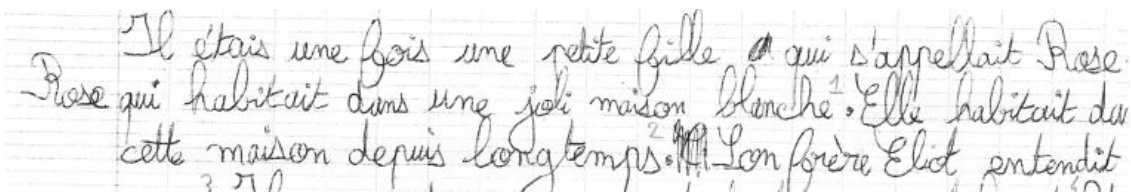
Étude des types de deuxième expansion dans la première mention du référent du SN démonstratif cette maison

Comme nous l'avons vu *supra,* le référent du SN démonstratif *cette maison* est majoritairement expansé à l'aide d'une seule expansion. Dans le corpus, il n'y a que sept textes dans lesquels le SN référant au démonstratif *cette maison* est expansé par deux modifieurs, comme l'indique le tableau 26 :

Type de deuxième expansion de M1 pour *cette maison*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
syntagme adjectival	1	2%	0	0%	2	3%	**3**	2%
syntagme prépositionnel	1	2%	3	5%	1	2%	**4**	2%
non pertinent	62	97%	53	95%	62	95%	178	96%
Total	64	100%	56	100%	65	100%	185	100%

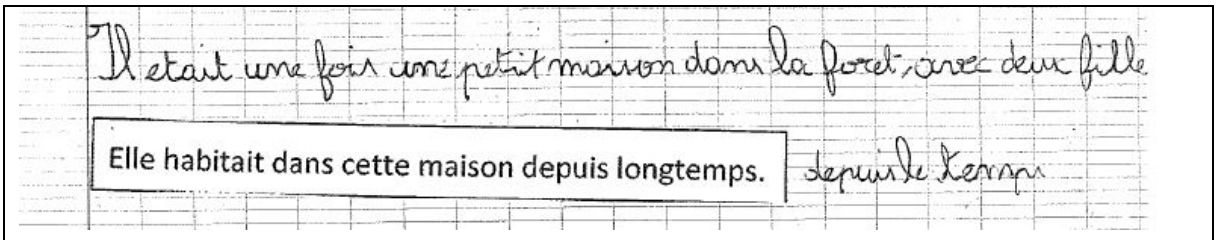
Tableau 26 : Type de deuxième expansion dans les formes de la première mention du référent du SN démonstratif *cette maison*

Les types de deuxième expansion sont soit des syntagmes adjectivaux (3 textes) comme dans (82) avec une *jolie maison* ***blanche,*** soit des syntagmes prépositionnels (5 textes), comme dans (83) avec *une petite maison* ***dans la forêt*** :


Il étais une fois une petite fille qui s'appellait Rose.
Rose qui habitait dans une joli maison blanche. Elle habitait dans
cette maison depuis longtemps. Son frère Eliot entendit

Il était une fois une petite fille qui s'appelait Rose qui habitait dans une jolie maison blanche. *Elle habitait dans cette maison depuis longtemps.*

Figure 82 : Texte d'Olivia, CE2 U


Il etait une fois une petit maison dans la foret, avec deux fille

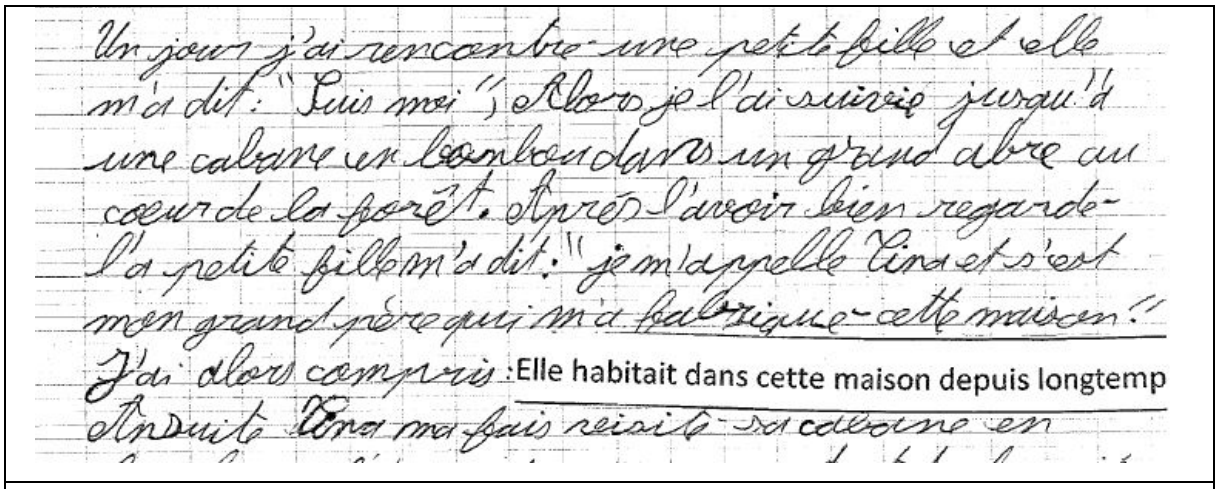
Elle habitait dans cette maison depuis longtemps.

depuis le temps

Il était une fois une petite maison dans la forêt avec deux filles. *Elle habitait dans cette maison depuis longtemps.*

Figure 83 : Texte de Soliana, CM1 T1

Enfin, dans l'ensemble du corpus, un seul texte comporte une première mention expansée par trois modifieurs. La troisième expansion est le syntagme prépositionnel *au cœur de la forêt* dans (84) :


Un jour j'ai rencontré une petite fille et elle m'a dit : "Suis moi", Alors je l'ai suivie jusqu'à une cabane en bambou dans un grand arbre au coeur de la forêt. Après l'avoir bien regardé-l'a petite fille m'a dit : "je m'appelle Tina et c'est mon grand père qui m'a fabrique cette maison !" J'ai alors compris : Elle habitait dans cette maison depuis longtemp

Un jour j'ai rencontré une petite fille et elle m'a dit : « Suis-moi ». Alors je l'ai suivie jusqu'à une cabane en bambou dans un grand arbre au cœur de la forêt. Après l'avoir bien regardée, la petite fille m'a dit : « je m'appelle Tina et c'est mon grand-père qui m'a fabriqué cette maison ». J'ai alors compris : *Elle habitait dans cette maison depuis longtemps.*

Figure 84 : Texte de Naélia, CM1 T

10.1.3. Etude des fonctions syntaxiques de la première mention du référent du SN démonstratif *cette maison*

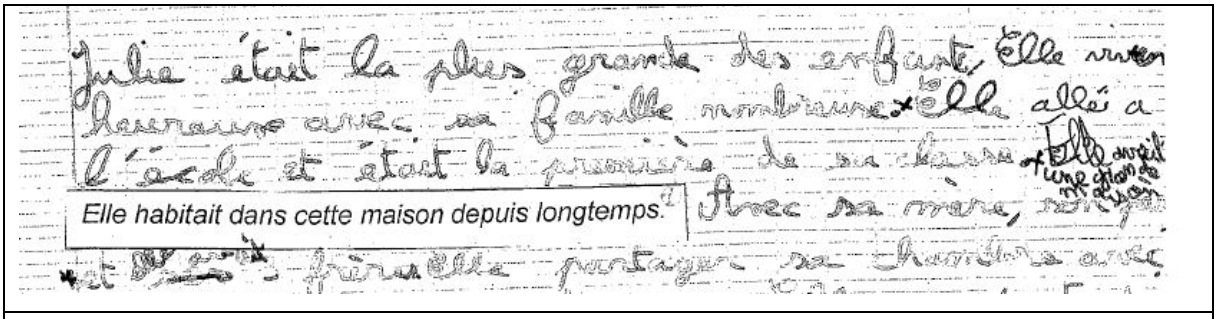
Lorsque la première mention d'un référent apparaît, c'est une information totalement nouvelle (cf. *statut de l'information*, chapitre 1, section 1.2.) qui représente le thème de l'énoncé (cf. *progression thématique*, chapitre 1, section1.2.3.). Dans ce cas, il est attendu que cette nouvelle information prenne la place de sujet syntaxique et se retrouve en position initiale dans l'énoncé.

Dans la tâche-problème de notre étude, le SN démonstratif *cette maison* apparaît dans la même phrase (P1) que le syntagme pronominal *elle*. Il y a donc dans la même phrase deux occurrences qui nécessitent une introduction préalable : un référent humain animé (le personnage féminin) et un référent non humain inanimé (la maison). Les caractéristiques différentes de ces deux entités ont une influence sur leur traitement. En effet, d'un point de vue sémantique, les caractéristiques des entités portées par les mots influent sur leur saillance (Lyons, 1980 ; Pattabhiraman, 1993 ; cf. chapitre 1, section 1.2.4.). Le caractère « humain » ou « animé » d'une entité augmente sa saillance, par rapport à une entité « non animée » ou un animal (Lyons, 1980). Ainsi, un apprenti scripteur ayant des difficultés à gérer plusieurs entités à la fois va avoir tendance à privilégier l'introduction du référent du pronom personnel *elle* plutôt que celle du référent du SN démonstratif *cette maison*. Et ce d'autant plus que dans P1, le pronom personnel est en position de sujet syntaxique et le SN démonstratif en position de complément de verbe. Si l'élève réussit à introduire le référent du SN démonstratif, il aura tendance à garder la même « hiérarchie » que celle de P1 (stratégie de non changement de rôle, cf. chapitre 3, section 3.1.) et le référent du SN démonstratif devrait alors se retrouver en position de complément de verbe, plus rarement en position de sujet syntaxique.

Cette hypothèse est confirmée par nos résultats : 44% des élèves du corpus emploient la première mention du référent du SN démonstratif *cette maison* en position de complément de verbe, comme l'illustrent le tableau 27 et l'exemple (85) ci-après.

Fonctions syntaxiques de M1 pour *cette maison*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
complément de verbe	29	45%	20	36%	33	51%	82	**44%**
complément de phrase	6	9%	5	9%	7	11%	18	**10%**
sujet	4	6%	5	9%	5	8%	**14**	8%
locution prépositionnelle	0	0%	1	2%	0	0%	**1**	1%
construction présentative	4	6%	4	7%	0	0%	**8**	4%
cas d'inférence	3	5%	3	5%	6	9%	**12**	6%
P1 = seule mention	18	28%	18	32%	13	20%	49	26%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 27 : Différents types de fonctions syntaxiques prises par la première mention du référent du SN démonstratif *cette maison*



Julie était la plus grande des enfants. Elle vivait heureuse avec sa famille nombreuse. Elle allait à l'école et était la première de sa classe. Elle avait une grande maison. *Elle habitait dans cette maison depuis longtemps.*

Figure 85 : Texte de Grégoire, CM2 B

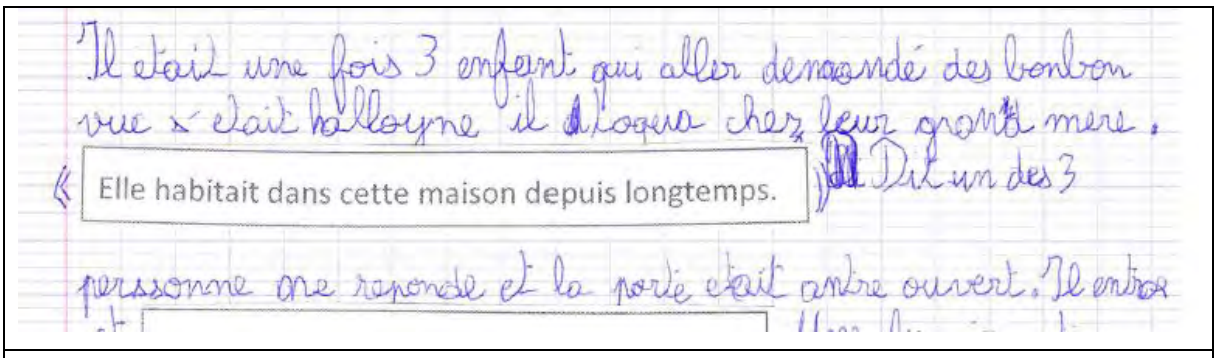
Dans cet exemple, les référents du pronom personnel *elle* et du SN démonstratifs *cette maison* sont dans les mêmes positions syntaxiques que dans P1 : le nom propre *Julie* et les pronoms *elle* sont en position de sujet syntaxique et le SN indéfini *une grande maison* en position de complément de verbe. Dans cet exemple également, la source du SN démonstratif *cette maison* apparaît dans la phrase précédant P1 : *Elle avait une grande maison*. La phrase dans laquelle apparaît le référent du SN démonstratif *cette maison* constitue un ajout de la part de l'élève, dans un second temps d'écriture. Cet ajout local différé montre que l'élève a réalisé, sans doute après relecture, que le référent du SN démonstratif *cette maison* n'avait pas été introduit.

La position de complément de verbe fait du SN démonstratif le rhème de l'énoncé. Si l'on a affaire à une progression linéaire, on s'attend à ce qu'il devienne le thème de l'énoncé suivant. Or il n'en est rien : dans P1, qui est la phrase suivante, le thème de l'énoncé est le pronom personnel *elle* et non le SN démonstratif *cette maison*. L'occurrence *maison* garde donc un aspect « secondaire » par rapport au personnage féminin.

De manière plus marginale, on constate que la première mention du référent du SN démonstratif *cette maison* apparaît dans diverses autres positions syntaxiques, témoignant d'une construction plus ou moins locale de l'entité :

- complément de phrase (10% des textes) : il s'agit de formes telles que *elle vivait à la campagne, dans une toute petite maison* (Léa, CM2 S).
- sujet (14 textes) : il s'agit de formes telles que *son ancienne maison lui manque énormément* (Cindy, CM1 T1).
- construction présentative (7 textes) : *Il était une fois une maison très jolie* (Chemsa, CE2 U)
- locution prépositionnelle (1 texte) : *elle sortit en dehors de la maison* (Amine, CM1 B)

Dans le tableau 27 *supra* des fonctions syntaxiques de la première mention du référent du SN démonstratif *cette* maison, la catégorie *cas d'inférence* représente 12 textes du corpus et correspond à des textes dans lesquels il n'y a pas référence à un terme explicitement mentionné dans l'énoncé. Dans ces textes, la référence au SN démonstratif *cette maison* entraîne une inférence, comme dans l'exemple (86) :



Il était une fois 3 enfants qui allaient demander des bonbons vu que c'était Halloween, il toqua chez leur grand-mère. « *Elle habitait dans cette maison depuis longtemps.* » dit un des 3. Personne ne répondait et la porte était entrouverte.

Figure 86 : Texte de Jonas, CE2 T

Dans la proposition *il toqua chez leur grand-mère*, l'inférence est portée par le verbe *toquer* car on comprend sans peine que si l'on toque quelque part, c'est à la porte d'une maison. En outre, le syntagme prépositionnel *chez leur grand-mère* confirme qu'il s'agit bien d'une maison. Ainsi, le référent du SN démonstratif *cette maison* est introduit sans mention explicite mais à l'aide d'une inférence préalable à P1.

10.2. Maintien du référent du SN démonstratif *cette maison*

Après avoir introduit le référent, l'élève peut choisir de le maintenir tout au long de son discours, bien qu'en tant qu'indication spatiale, on ne s'attend pas à ce que l'occurrence *maison* soit maintenue très longtemps. Nous allons donc étudier à présent comment la référence au SN démonstratif *cette maison* est maintenue dans les textes. En 10.2.1., nous dressons une vue générale du maintien de la référence à travers l'étude du nombre de mentions des CDR et de leur longueur moyenne. En 10.2.2., nous étudions de manière plus précise différents cas de figure, selon qu'il s'agit d'anaphore (1 mention + P1 soit 2 mentions au total), de CDR minimales (2 mentions + P1 soit 3 mentions au total) ou de CDR plus longues (3 mentions ou plus + P1 soit 4 mentions ou plus au total).

10.2.1. Vue d'ensemble des CDR incluant le SN démonstratif *cette maison*

De manière générale, l'étude du nombre de mentions (ou maillons) de la CDR incluant le SN démonstratif *cette maison* montre que les CDR les plus longues comportent 7 maillons + celui compris dans P1, soit 8 maillons au total, comme l'illustre le tableau 28. On constate également que les CDR les plus longues sont représentées dans chacun des niveaux de classe mais par un seul texte à chaque fois.

Nombre de mentions dans la CDR de *cette maison*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
1 mention + P1	21	33%	16	29%	21	32%	58	**31%**
2 mentions + P1	18	28%	11	20%	17	26%	46	25%
3 mentions + P1	5	8%	6	11%	7	11%	18	10%
4 mentions + P1	0	0%	4	7%	4	6%	8	4%
5 mentions + P1	1	2%	0	0%	1	2%	2	1%
6 mentions + P1	0	0%	0	0%	1	2%	1	1%
7 mentions + P1	**1**	2%	**1**	2%	**1**	2%	3	2%
P1 = seule mention	18	28%	18	32%	13	20%	49	26%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 28 : Nombre de mentions dans la CDR incluant le référent du SN démonstratif *cette maison*

Comme nous l'avons dit *supra*, le SN démonstratif *cette maison* est une indication spatiale contenue dans P1. À ce titre, on ne s'attend pas à ce que l'occurrence *maison* fasse l'objet d'une véritable CDR. La lecture du tableau ci-dessus confirme notre prédiction : plus les CDR comportent de maillons, moins elles sont nombreuses. Les cas d'anaphores (1 mention + P1 soit 2 mentions au total) sont d'ailleurs les plus représentés : 31% dans l'ensemble du corpus. La longueur moyenne des CDR incluant le SN démonstratif *cette maison* est en accord avec ce résultat. S'il augmente légèrement en fonction du niveau scolaire, le nombre moyen de maillons par CDR est d'environ 2 :

- 2,3 maillons au CE2
- 2,4 maillons au CM1
- 2,7 maillons au CM2

Le faible nombre de maillons peut être expliqué par le fait que l'occurrence *maison* est une entité non humaine et non animée. Dans les phrases imposées de la consigne, non seulement elle se retrouve en compétition avec d'autres entités que l'élève doit également gérer, mais surtout, elle se retrouve en compétition avec les entités humaines que sont les personnages humains de la consigne, représentés par les pronoms personnels *elle* et *il* ainsi que le SN *les enfants*. L'occurrence *maison* passe donc au second plan (Lyons, 1980 ; cf. chapitre 1, section 1.2.4.) et les élèves ne s'attardent pas sur elle, mis à part dans quelques rares textes dont c'est le sujet principal de l'histoire.

10.2.2. Différentes configurations des CDR

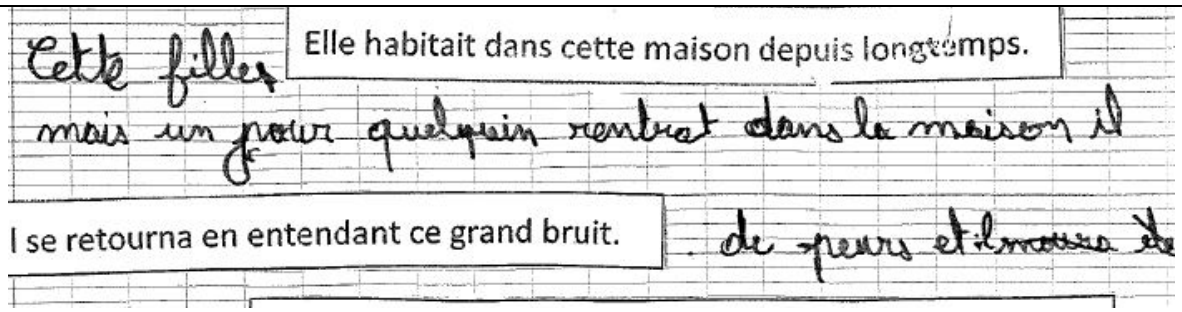
Nous l'avons vu, notre corpus regroupe plusieurs configurations de CDR. Il peut s'agir d'anaphore ou de cataphore (1 mention + P1), de CDR minimales (2 mentions + P1) ou de CDR plus longues (3 mentions ou plus + P1). Chacune de ces configurations fait l'objet d'une étude plus précise.

Étude des cas d'anaphore ou de cataphore : 1 mention + P1

Le cas de figure le plus fréquent dans notre corpus lorsqu'il s'agit de maintenir la référence à l'occurrence *maison* est le recours à la configuration suivante : 1 mention + P1, soit deux mentions au total (cf. tableau 28 *supra*). Dans la configuration « 1 mention + P1 », on trouve deux types de configurations : les cas d'anaphore et les cas de cataphore.

On parle de cas d'anaphore lorsqu'il y a une mention du référent avant P1 et celle déjà présente dans P1 (Reichler-Béguelin, 1988a ; cf. chapitre 2). Les caractéristiques de la première mention du référent sont alors les mêmes que celles analysées dans la section 10.1.1. lorsque nous avons étudié l'introduction du référent : la forme linguistique la plus fréquente est le SN indéfini en position de complément de verbe. Les cas d'anaphore témoignent d'une gestion de la référence au SN démonstratif plutôt locale ou « en arrière-plan ». En effet, le petit nombre de mentions du référent ne permet pas d'en faire une entité véritablement présente dans le discours ou bien de construire une histoire entière dont elle serait le sujet thématique.

On parle de cataphore (cf. chapitre 2, section 2.1.) lorsque le SN démonstratif *cette maison* de P1 est la première mention du référent. La source du référent est alors localisée après P1, comme dans (87) avec le SN défini *la maison*.



Cette fille, *elle habitait dans cette maison depuis longtemps.* Mais un jour, quelqu'un rentra dans la maison. *Il se retourna en entendant ce grand bruit*.

Figure 87 : Texte de Carl, CM1 T1

La première mention du référent étant situé après P1, on s'attend à ce qu'elle apparaisse principalement sous la forme d'un SN défini, contrairement aux cas d'anaphore, car le référent est devenu accessible (cf. chapitre 1, section 1.2.). Le tableau 29 montre que c'est effectivement le cas le plus fréquent : 53% des cas de cataphore.

Formes de M1 dans les cas de cataphore								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN défini	3	43%	5	100%	0	0%	8	**53%**
SN possessif	1	14%	0	0%	1	33%	2	13%
SN pronominal	1	14%	0	0%	0	0%	1	7%
SN indéfini	1	14%	0	0%	1	33%	2	13%
autre	1	14%	0	0%	1	33%	2	13%
Total	7	100%	5	100%	3	100%	15	100%

Tableau 29 : Formes de la première mention du référent du SN démonstratif *cette maison* dans les cas de cataphore (P1 + 1 mention)

En outre, l'étude des fonctions syntaxiques de la première mention du référent du SN démonstratif *cette maison* dans les cas d'anaphore ou de cataphore montre que la position la plus fréquente est celle de complément de verbe, quelle que soit la localisation de la source (avant ou après P1).

Étude des CDR minimales : 2 mentions + P1

Le deuxième cas de figure le plus fréquent dans notre corpus lorsqu'il s'agit de maintenir la référence à l'occurrence *maison* est le recours à la configuration suivante : 2

mention + P1 (cf. tableau 28 *supra*), c'est-à-dire 3 mentions au total. Il s'agit là de la configuration minimale dans laquelle on peut parler de CDR (Schnedecker, 1997 ; cf. chapitre 2, section 2.2.). On parle donc de CDR minimale, en opposition avec les CDR contenant plus de 3 maillons que nous aborderons plus tard.

Les CDR minimales témoignent d'une gestion de la référence plus globale. En effet, en mentionnant le référent du SN démonstratif *cette maison* plusieurs fois en dehors de P1, l'élève commence à construire son personnage (en l'occurrence l'entité *maison*) en lui octroyant une place plus importante. Encore une fois, le caractère non humain et non animé de l'occurrence *maison* ne lui permet pas *a priori* d'avoir la même importance que les personnages humains de la consigne. Néanmoins, étant mentionné plusieurs fois, l'occurrence *maison* devient un élément secondaire qui garde une relative importance, comme un élément du décor par exemple.

Si l'étude de la première mention du référent du SN démonstratif *cette maison* révèle que la forme la plus fréquente de la première mention est le SN indéfini, l'étude de la deuxième mention du référent dans les CDR minimales montre que la forme la plus fréquente est le SN défini, comme l'illustre le tableau 30 : 52% dans l'ensemble des CDR minimales.

Formes de M2 dans les CDR minimales								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN défini	8	44%	7	64%	9	53%	24	**52%**
SN possessif	2	11%	2	18%	1	6%	5	11%
SN démonstratif	4	22%	0	0%	1	6%	5	11%
SN indéfini	0	0%	0	0%	2	12%	2	4%
autre	4	22%	2	18%	4	24%	10	22%
Total	18	100%	11	100%	17	100%	46	100%

Tableau 30 : Formes de la deuxième mention du référent du SN démonstratif *cette maison* dans les CDR minimales

La prédominance du SN défini en deuxième mention est justifiée par le fait que le référent étant déjà introduit au préalable (première mention), il devient accessible pour le lecteur. Le SN défini est une forme de moyenne accessibilité (Ariel, 1980 ; cf. chapitre 1, section 1.2.2.) qui a toute sa place en deuxième mention. En effet, les travaux de Landragin & *al.* (2015 ; cf. chapitre 2, section 2.2.4.) notamment, précisent qu'une chaîne coréférentielle prototypique

d'un référent peut se retrouver sous la forme : nom propre/SN indéfini – description définie – successions de pronoms personnels de troisième personne.

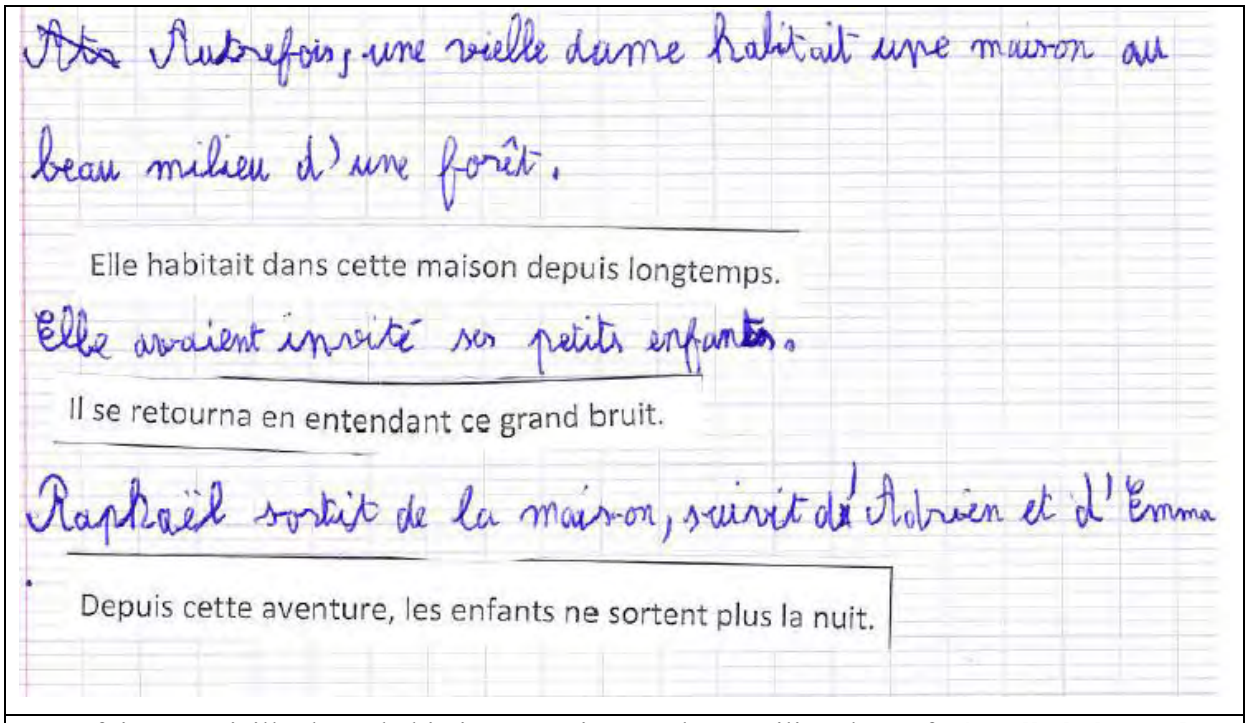
L'étude des fonctions syntaxiques de la deuxième mention des CDR minimales (cf. tableau 31) montre que la position de complément de verbe est la fonction la plus fréquente : 65% dans l'ensemble des CDR minimales.

Fonctions syntaxiques de M2 dans les CDR minimales								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
complément de verbe	12	67%	7	64%	11	65%	30	**65%**
complément de phrase	3	17%	2	18%	2	12%	7	15%
complément du nom	0	0%	1	9%	0	0%	1	2%
sujet	3	17%	0	0%	2	12%	5	11%
phrase nominale	0	0%	0	0%	2	12%	2	4%
autre	0	0%	1	9%	0	0%	1	2%
Total	18	100%	11	100%	17	100%	46	100%

Tableau 31 : Fonctions syntaxiques de la deuxième mention du référent du SN démonstratif *cette maison* dans les CDR minimales

Comme nous l'avons vu en détail dans l'étude des fonctions syntaxiques de la première mention du référent (cf. ce chapitre, section 10.1.1.3.), l'élève a tendance à garder la même « hiérarchie » que celle de P1 et le référent du SN démonstratif se retrouve en position de complément de verbe, plus rarement en position de sujet syntaxique (stratégie de non changement de rôle).

La deuxième mention du référent sous la forme d'un SN défini en position de complément de verbe se retrouve dans (88) :


Autrefois, une vieille dame habitait une maison au beau milieu d'une forêt.

Elle habitait dans cette maison depuis longtemps.

Elle avaient invité ses petits enfants.

Il se retourna en entendant ce grand bruit.

Raphaël sortit de la maison, suivit d'Adrien et d'Emma

Depuis cette aventure, les enfants ne sortent plus la nuit.

Autrefois, une vieille dame habitait une maison au beau milieu d'une forêt. *Elle habitait dans cette maison depuis longtemps.* Elle avait invité ses petits-enfants. *Il se retourna en entendant ce grand bruit.* Raphaël sortit de la maison, suivi d'Adrien et d'Emma. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 88 : Texte d'Alban, CE2 P

Dans cet exemple, la CDR incluant le SN démonstratif *cette maison* comporte deux maillons. La première mention du référent apparaît sous la forme du SN indéfini *une maison* en position de complément de verbe. La deuxième mention du référent apparait quelques lignes plus loin sous la forme du SN défini *la maison* en position de complément de verbe. La cohérence textuelle est respectée.

Étude des CDR plus longues : 3 mentions (ou plus) + P1

Les textes présentant des CDR plus longues, c'est-à-dire 3 mentions (ou plus) + P1 sont moins nombreux : entre 8 textes et un seul dans l'ensemble du corpus. Les textes comportant les CDR les plus longues, c'est-à-dire qui comportent 7 mentions + P1 (soit 8 mentions au total), sont minoritaires : 3 textes dans l'ensemble du corpus. On en trouve un exemplaire dans chaque niveau de classe. Nous ne pourrions ici faire une revue détaillée de chacune des mentions du référent du SN démonstratif *cette maison* comme nous l'avons fait

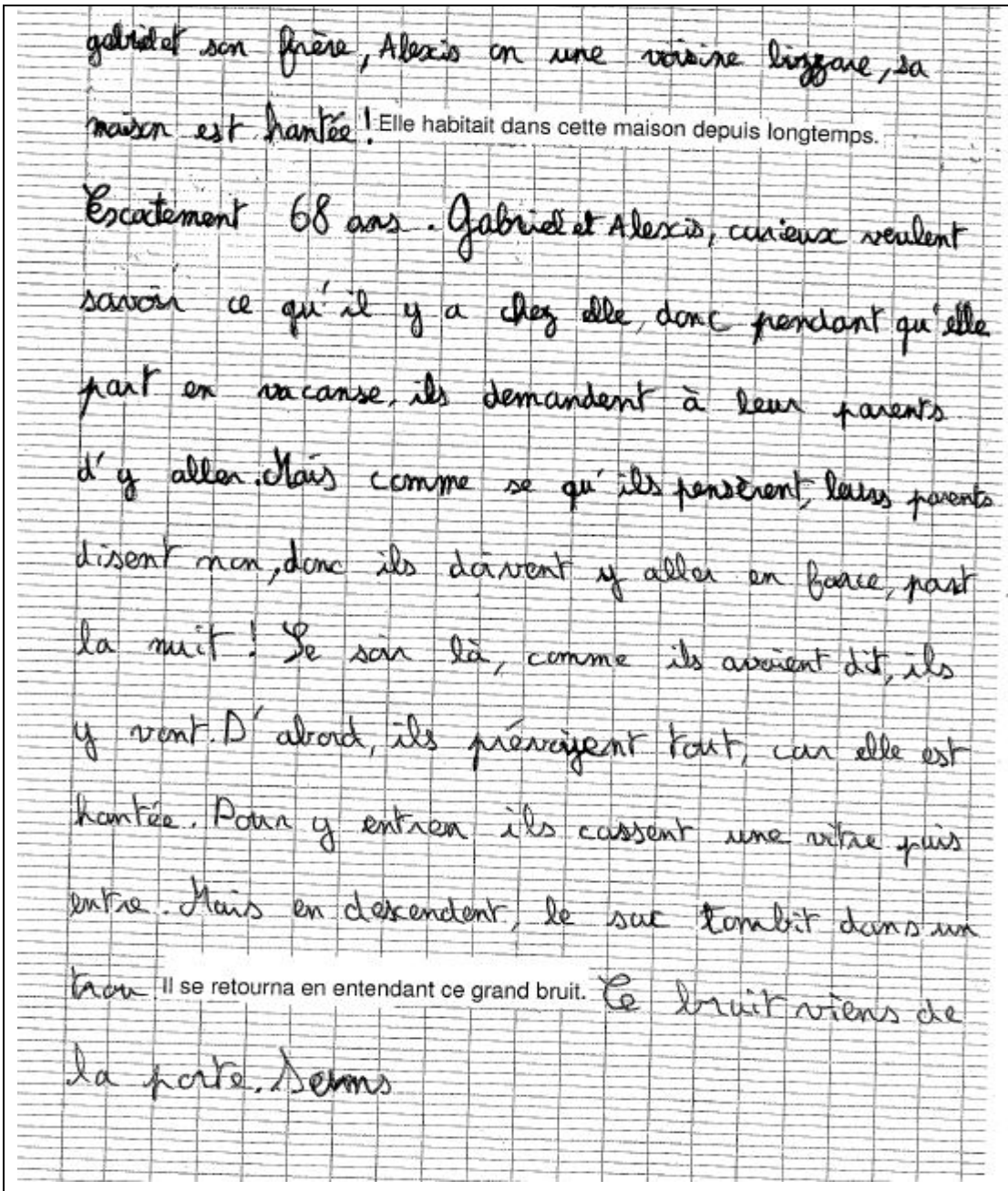
jusqu'à présent. Nous nous contenterons donc d'une vue générale des compositions des CDR comportant 3 maillons (ou plus) + P1.

Quel que soit le nombre de maillons de la CDR incluant le SN démonstratif *cette maison*, le SN défini est la forme de reprise la plus fréquente. De prime abord, ce résultat peut paraître différent de ceux obtenus dans divers travaux qui s'intéressent au maintien de la référence chez les enfants ou les adultes. En effet, Karmiloff-Smith (1980 ; 1981 ; cf. chapitre 3, section 3.1.2.2.) par exemple, montre que le pronom est la forme privilégiée de la reprise. Cependant, dans notre tâche-problème, l'occurrence *maison* est un élément secondaire de l'histoire, contrairement aux personnages humains qui en sont les protagonistes. Les travaux d'Hickmann (1980, 1984, 1987 ; cf. chapitre 3, section 3.1.2.2.) par exemple, ont ainsi mis en évidence le fait que les enfants utilisent plutôt des syntagmes nominaux et quelques pronoms pour référer aux personnages secondaires. Le recours au SN défini pour maintenir la référence aux entités secondaires du récit, en l'occurrence ici au SN démonstratif *cette maison*, est donc en accord avec ces travaux.

En ce qui concerne les expansions, on constate des formes expansées jusqu'à la 4e mention du référent du SN démonstratif *cette maison,* mais pas au-delà. Les 3e et 4e mentions sont principalement expansées à l'aide d'un syntagme adjectival.

Enfin, l'étude des fonctions syntaxiques des différentes mentions montre qu'elles sont exclusivement en position de complément, le complément de verbe étant la position syntaxique la plus fréquente, à l'image de la position syntaxique occupée par le SN démonstratif *cette maison* dans P1.

Nous retrouvons ces éléments dans le texte de Mathieu (CM2 S), inachevé, que nous reproduisons dans son intégralité et qui comporte la plus longue CDR incluant le SN démonstratif *cette maison* dans notre corpus, puisqu'elle comporte 7 mentions du référent + P1.



gabriel et son frère, Alexis on une voisine bizzare, sa maison est hantée ! Elle habitait dans cette maison depuis longtemps.
Exactement 68 ans. Gabriel et Alexis, curieux veulent savoir ce qu'il y a chez elle, donc pendant qu'elle part en vacanse, ils demandent à leur parents d'y aller. Mais comme se qu'ils pensèrent, leurs parents disent non, donc ils doivent y aller en force, part la nuit ! Se soir là, comme ils avaient dit, ils y vont. D'abord, ils prévoyent tout, car elle est hantée. Pour y entrer ils cassent une vitre puis entre. Mais en descendent, le sac tombit dans un trou. Il se retourna en entendant ce grand bruit. Ce bruit viens de la porte. Demms

Gabriel et son frère, Alexis ont une voisine bizarre, sa maison est hantée ! *Elle habitait dans cette maison depuis longtemps.* Exactement 68 ans. Gabriel et Alexis, curieux, veulent savoir ce qu'il y a chez elle, donc pendant qu'elle part en vacances, ils demandent à leurs parents d'y aller. Mais comme ce qu'ils pensèrent, leurs parents disent « non », donc ils doivent y aller en force, par la nuit ! Ce soir-là, comme ils avaient dit, ils y vont. D'abord, ils prévoient tout, car elle est hantée. Pour y entrer, ils cassent une vitre puis entrent. Mais en descendant, le sac tomba dans un trou. *Il se retourna en entendant ce grand bruit.* Ce bruit vient de la porte.

Figure 89 : Texte de Mathieu, CM2 S

Dans ce texte, le référent au SN démonstratif *cette maison* est mentionné une seule fois avant P1, sous la forme du SN possessif *sa maison*. La forme employée n'est pas la forme la plus attendue. En effet, en première mention, on s'attend à trouver un SN indéfini étant donné que l'entité est entièrement nouvelle et de fait, peu accessible. Toutefois, par le recours au SN possessif, l'élève choisi de « raccrocher » cette information au personnage féminin de la voisine, personnage qu'il a présenté au préalable.

Toutes les autres mentions du référent apparaissent dans le texte après P1. Mis à part le syntagme prépositionnel *chez elle*, toutes les formes de reprise sont des pronoms, conformément à ce qui pourrait être attendu étant donné que le référent a maintenant une forte accessibilité. L'occurrence *maison*, même si elle n'est que deux fois en position de sujet syntaxique dans le texte, reste dans le focus d'attention du lecteur.

10.3. Bilan de l'étude du référent du SN démonstratif *cette maison*

De manière générale, l'étude des formes de la première mention du référent du SN démonstratif *cette maison* montre que :

- la forme privilégiée pour introduire le référent est le SN indéfini, avec une fréquence d'emploi plus forte au CM2.
- la première mention du référent apparaît le plus fréquemment en position de complément de verbe quel que soit le niveau de classe, plus rarement en position de sujet syntaxique.
- les formes lexicales des SN employés en première mention sont majoritairement le SN *maison*. On trouve toutefois quelques termes spécifiques tels que *manoir, château* ou *cabane* dans les niveaux CM1 et CM2.
- lorsque la forme de la première mention est expansée, elle l'est majoritairement à l'aide d'une seule expansion. On constate également que plus le niveau augmente, plus la fréquence d'emploi d'une expansion augmente.
- les types d'expansions utilisées sont principalement des syntagmes adjectivaux et dans une moindre mesure, des syntagmes prépositionnels. L'utilisation de syntagmes adjectivaux croît lorsque le niveau de classe augmente.

L'étude du maintien de la référence au SN démonstratif *cette maison* montre plusieurs manières de « gérer » la référence. La première est une gestion plutôt locale du référent : les textes présentent deux mentions du référent, c'est-à-dire 1 mention + celle de P1. C'est d'ailleurs la longueur moyenne des CDR dans notre corpus. Dans cette configuration, la mention employée par les élèves (ou source) est localisée soit avant P1 (anaphore), soit après P1 (cataphore). Selon la localisation de la source, la forme de la première mention change. Dans les cas d'anaphore, la forme la plus fréquente est le SN indéfini. Dans les cas de cataphore, c'est le SN défini. Cependant, quelle que soit la localisation de la source, la fonction syntaxique la plus fréquemment utilisée est le complément de verbe.

La deuxième manière de « gérer » la référence se traduit par une gestion plus globale du référent : les élèves le mentionnent plusieurs fois dans leur texte, ce qui constitue des CDR plus ou moins longues. Lorsque la CDR déployée pour le SN démonstratif *cette maison* est minimale (2 mentions + P1, soit 3 mentions au total), ou bien plus longue (3 mentions ou plus + P1), la forme de la deuxième mention est majoritairement un SN défini, en position de complément de verbe. Dans les CDR les plus longues de notre corpus, qui peuvent atteindre jusqu'à 7 mentions + P1 (soit 8 mentions au total), on constate une prédominance de SN définis et de pronoms comme formes de reprise du référent.

CHAPITRE 11 : Introduction et maintien du référent du SN démonstratif *ce grand bruit*

Dans la tâche-problème que doivent réaliser les élèves de notre étude, la deuxième phrase à insérer (P2) est la suivante :

Il se retourna en entendant ce grand bruit.

Dans cette phrase, l'emploi du SN démonstratif *ce grand bruit* demande un traitement particulier de la part des élèves, pour deux raisons. La première tient au fait que les SN démonstratifs, comme nous l'avons vu dans le chapitre précédent, sont des expressions qui nécessitent le recours à une entité identifiable venant d'être évoquée dans le discours au moyen d'expressions linguistiques (Charolles, 2002 ; cf. chapitre 1, section 1.4.2.1.). L'emploi d'un SN démonstratif induit la présence d'une source avec laquelle il peut être mis en relation (Marandin, 1986 : 77).

La seconde raison est en lien avec le sémantisme du nom *bruit*. En tant que mot « plein » (Mortureux, 1998), le nom *bruit* possède une certaine valeur dénominative qui lui permet d'évoquer la réalité qui lui est associée. Mais le terme *bruit* est générique : selon la définition du Tlf, c'est « un ensemble de sons ». Un bruit peut donc avoir une multitude de causes. De ce fait, on s'attend à ce que le référent du SN démonstratif *ce grand* bruit se réalise sous des formes linguistiques variées : onomatopées, SN, phrase complète qui nécessite une inférence… En outre, dans P2, le SN *bruit* est expansé par le syntagme adjectival *grand* (*grand bruit)*. L'expansion choisie confère au nom *bruit* une propriété particulière qui doit être prise en compte : un *grand* bruit n'est pas n'importe quelle sorte de bruit. La résolution du SN démonstratif anaphorique *ce grand bruit* pose donc une question de cohérence textuelle essentielle qui implique de gérer à la fois l'introduction du SN démonstratif par un terme linguistique approprié et la cohésion sémantique du terme choisi en accord avec le SN *grand bruit*.

En 11.1, nous allons étudier la manière dont les élèves introduisent le référent du SN démonstratif *ce grand bruit*, en analysant à la fois les formes de la première mention du référent, leurs éventuelles expansions et leurs fonctions syntaxiques. Après avoir introduit le référent, les élèves peuvent choisir de le maintenir tout au long du discours, ou tout du moins

dans une partie du discours. En 11.2., nous allons, de ce fait, étudier la manière dont les élèves maintiennent la référence tout au long du texte.

11.1. Introduction du référent du SN démonstratif *ce grand bruit*

De manière générale, dans notre corpus, on observe trois localisations possibles de l'expression référentielle qui introduit le référent (ou *source*), comme le montre le tableau 32 :

- la source (ou référent) est introduite avant la phrase imposée et la référence est maintenue après
- la source n'est pas introduite avant la phrase imposée mais elle est localisée après
- la source est absente : la seule mention du référent est celle de la phrase imposée

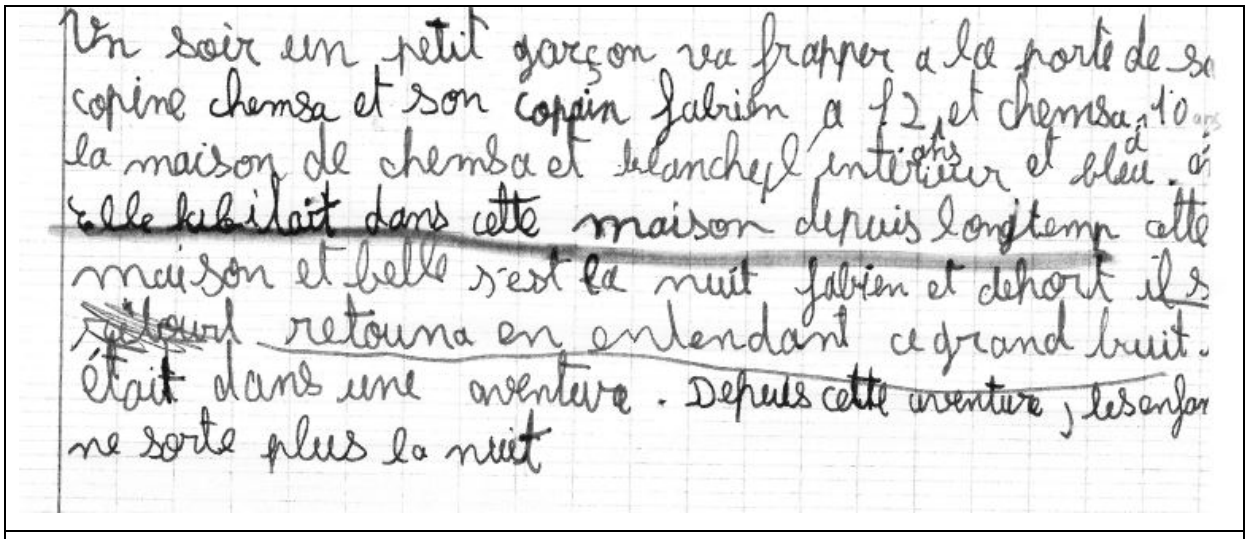
Localisation de la source du SN *ce grand bruit*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
avant P2 et maintien après	26	41%	22	39%	32	49%	80	**43%**
après P2	25	39%	11	20%	14	22%	50	**27%**
source absente	13	20%	23	41%	19	29%	55	**30%**
Total	64	100%	56	100%	65	100%	185	100%

Tableau 32 : Localisation de la source du SN démonstratif *ce grand bruit*

Sur l'ensemble du corpus, le référent est introduit avant P2 et maintenu par la suite dans 43% des textes. Globalement, les textes présentant cette configuration sont plus nombreux au CM2 que dans les autres niveaux de classe : 41% au CE2, 39% au CM1, 49% au CM2.

Dans 27% des textes du corpus, la source du SN démonstratif *ce grand bruit* est localisée après P2. Il s'agit de cataphores (cf. chapitre 2, section 2.1.1.1.) : le SN démonstratif *ce grand bruit* de P2 est la première mention du référent, repris par la suite sous diverses formes. Cette manière d'introduire le référent est plus fréquente lorsque les élèves sont plus jeunes : 39% au CE2, 20% au CM1 et 22% au CM2.

Enfin, sur l'ensemble du corpus, dans 30% des textes la source du SN démonstratif *ce grand bruit* est absente, comme dans (90) :


Un soir un petit garçon va frapper a la porte de sa copine chemsa et son copain fabien a 12 ans et chemsa a 10 ans la maison de chemsa et blanche l'intérieur est bleu. Elle habitait dans cette maison depuis longtemp cette maison et belle s'est la nuit fabien et dehort il se retourna en entendant ce grand bruit. était dans une aventure. Depuis cette aventure, les enfan ne sorte plus la nuit

Un soir un petit garçon va frapper à la porte de sa copine Chemsa et son copain Fabien a 12 ans et Chemsa a 10 ans. La maison de Chemsa est blanche, l'intérieur est bleu. *Elle habitait dans cette maison depuis longtemps.* Cette maison est belle. C'est la nuit, Fabien est dehors. *Il se retourna en entendant ce grand bruit.* (C') était dans une aventure. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 90 : Texte de Gabriel, CE2 U

On pourrait s'attendre à ce que la fréquence de cette absence décroisse avec l'âge dans la mesure où la maîtrise de l'introduction des référents est progressive (cf. chapitre 3). Or, on remarque que ce n'est pas le cas : c'est au CE2 que les textes dans lesquels P2 est la seule mention du référent sont les moins nombreux : 20% au CE2 contre 41% au CM1 et 29% au CM2. On remarque également le grand nombre de cas au CM1.

Après avoir dressé une vue d'ensemble des différentes manières de gérer la référence, nous allons étudier les formes de la première mention du référent du SN démonstratif *ce grand bruit.*

11.1.1. Etude des formes de la première mention du référent du SN démonstratif *ce grand bruit*

Le SN démonstratif *ce grand bruit* implique un haut degré d'accessibilité du référent. Pour que le lecteur comprenne à quelle entité réfère le SN démonstratif, l'introduction d'un référent préalablement à P2 s'impose. Le référent ainsi introduit constitue la première mention (M1) du SN démonstratif *ce grand bruit.*

Comme nous l'avons vu dans l'introduction de ce chapitre, le sémantisme du nom *bruit* contenu dans le SN démonstratif *ce grand bruit* implique que les mentions du référent se réalisent sous des formes linguistiques très variées, allant des formes « ordinaires » telles que SN (indéfini/défini/etc.) aux onomatopées, voire à des phrases entières qui expliquent la provenance du bruit, qu'elles impliquent une inférence ou non. Le tableau 33 reproduit les différents types de formes de la première mention et leur fréquence d'emploi dans notre corpus.

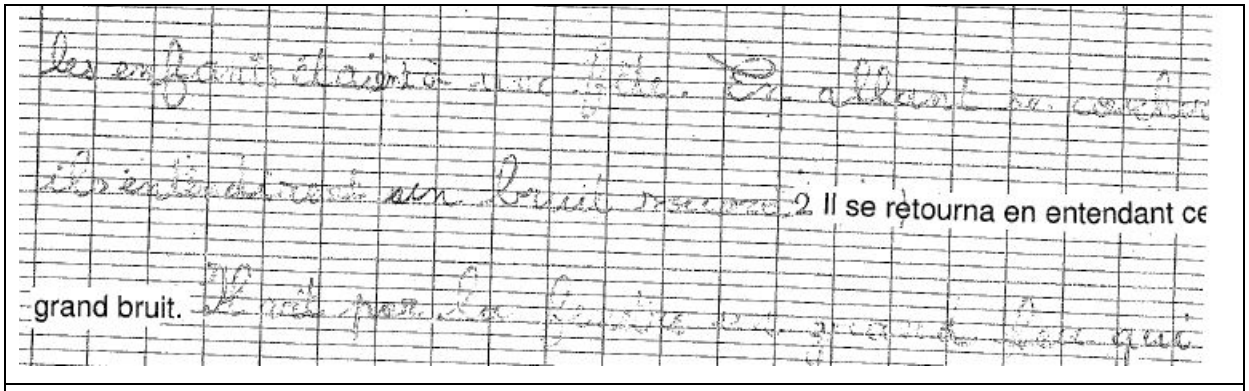
Formes de M1 pour *ce grand bruit*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN indéfini	16	25%	10	18%	16	25%	42	**23%**
onomatopée	5	8%	6	11%	6	9%	17	**9%**
phrase	24	38%	11	20%	19	29%	54	**29%**
SN défini	4	6%	2	4%	2	3%	**8**	4%
SN démonstratif	2	3%	3	5%	3	5%	**8**	4%
non pertinent	13	20%	24	43%	19	29%	56	30%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 33 : Formes de la première mention du référent du SN démonstratif *ce grand bruit*

Les différentes formes d'introduction du référent du SN démonstratif ce grand bruit

Comme nous l'avons déjà vu, les travaux qui se sont intéressés à l'introduction des référents dans un discours chez les enfants (Hickmann, 1984 ; 1987 par exemple, cf. chapitre 3, section 3.1.2.1.) ou les adultes (Givón, 1983 ; 1989 ; cf. chapitre 2, section 2.2.4.) montrent

que le SN indéfini est la forme privilégiée pour la première mention d'un référent. Toutefois dans notre corpus, ce n'est pas la forme la plus fréquente : elle ne représente que 23 % sur l'ensemble des textes, comme l'illustre l'exemple (91) :

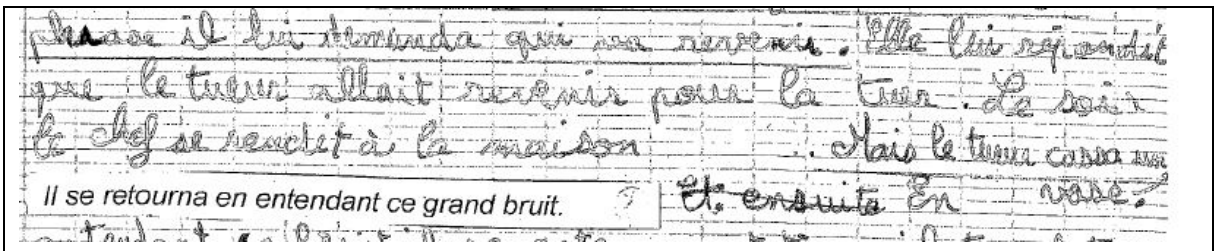


En allant se coucher ils entendirent un bruit sourd. *Il se retourna en entendant ce grand bruit.*

Figure 91 : Texte de Max, CM2 S

La lecture du tableau des formes de la première mention du référent du SN démonstratif *ce grand bruit* montre que 29% des textes du corpus introduisent le référent du SN démonstratif par une phrase entière. On constate également que cette forme est plus employée au CE2 qu'au CM2. On peut distinguer trois cas différents.

Dans le premier cas, la phrase ne comporte ni le mot *bruit,* ni le verbe *entendre,* ni aucun autre élément lexical dont le sémantisme à lui seul peut l'inclure dans le champ notionnel de *bruit*. Le lecteur doit réaliser un calcul interprétatif pour comprendre la provenance du bruit et le relier au SN démonstratif *ce grand bruit* de P2. C'est le cas par exemple dans (92) :



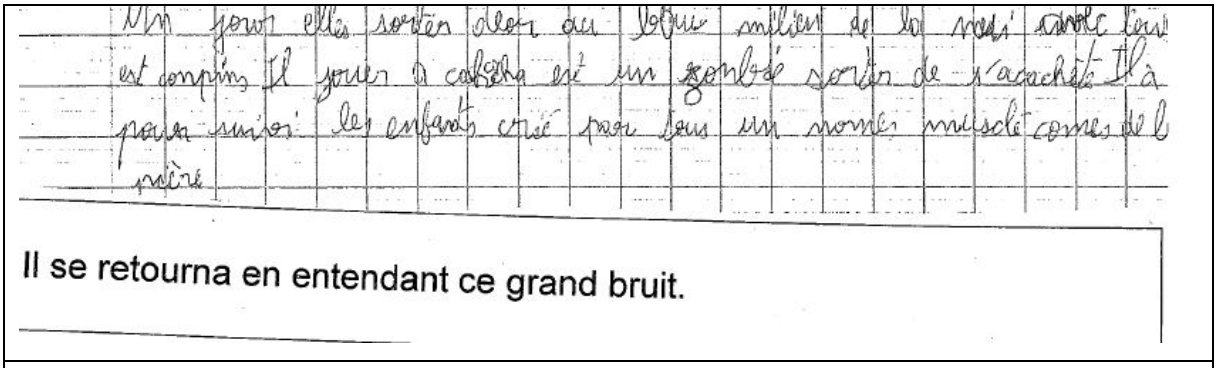
Elle lui répondit que le tueur allait revenir pour la tuer. Le soir, le chef se rendit à la maison. Mais le tueur cassa un vase. *Il se retourna en entendant ce grand bruit.*

Figure 92 : Texte d'Ingryd, CM2 B

Dans ce texte, la phrase *Mais le tueur cassa un vase* précède P2. Le lecteur doit procéder à un rapide calcul inférentiel pour comprendre que le fait de casser un vase produit un bruit et que c'est donc la source du bruit mentionné dans P2. Par ailleurs, l'action se passe le soir, on peut

donc raisonnablement penser qu'elle a lieu pendant la nuit (à moins bien sûr d'être en été dans un pays comme l'Islande ou le jour ne se couche quasiment jamais). Le bruit produit par le vase en se brisant peut être amplifié par le relatif silence de la nuit. L'élève satisfait ainsi également à la contrainte sémantique portant sur l'adjectif *grand*.

Dans le deuxième cas, la phrase ne contient ni le mot *bruit*, ni le verbe *entendre* mais elle comporte un élément lexical dont le sémantisme permet d'inférer directement l'existence du bruit. C'est le cas par exemple dans l'extrait (93), avec le verbe *crier* dans *les enfants criaient partout* :

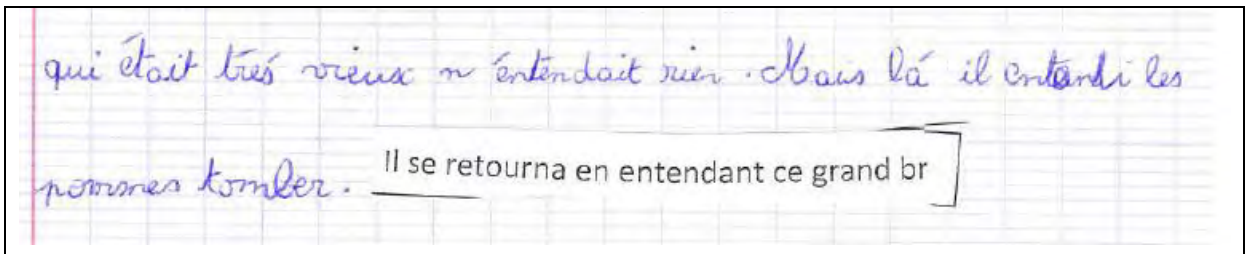

Un jour elles sorter dehors au beau milieu de la nuit avec tous
est compins Il jouer à cachacha est un zombie sorter de s'acachete Il à
pour suivi les enfants crié par tous un homme musclé comme de la
pierre

Il se retourna en entendant ce grand bruit.

Un jour elle sortit dehors au beau milieu de la nuit avec tous ses copains. Ils jouaient à cache-cache et un zombie sortit de sa cachette. Il a poursuivi les enfants (les enfants) criaient partout. Un homme musclé comme de la pierre, *il se retourna en entendant ce grand bruit.*

Figure 93 : Texte d'Anass, CM1 B

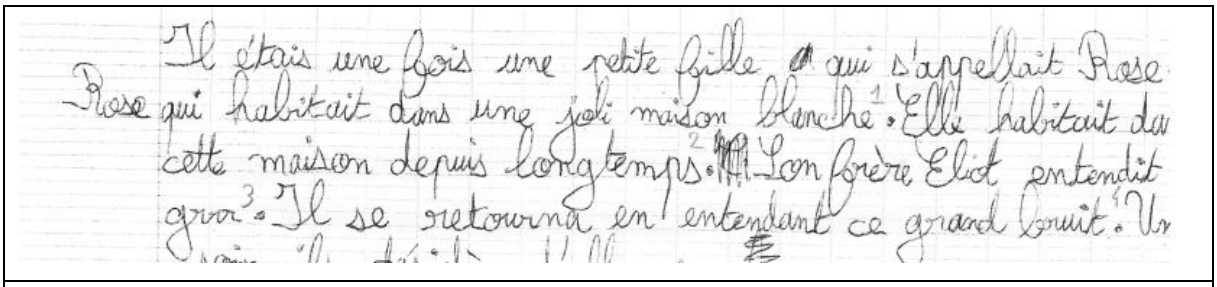
Dans le troisième cas, la phrase contient le verbe *entendre,* comme dans (94). Le verbe *entendre* porte en lui-même la notion de bruit puisque si on entend c'est qu'il y a un bruit de nature quelconque. L'inférence que le lecteur doit effectuer est donc moins coûteuse.


qui était trés vieux n'entendait rien. Mais lá il entendi les
pommes tomber. Il se retourna en entendant ce grand br

Mais là il entendit les pommes tomber. *Il se retourna en entendant ce grand bruit.*

Figure 94 : Texte de Naïla, CE2 P

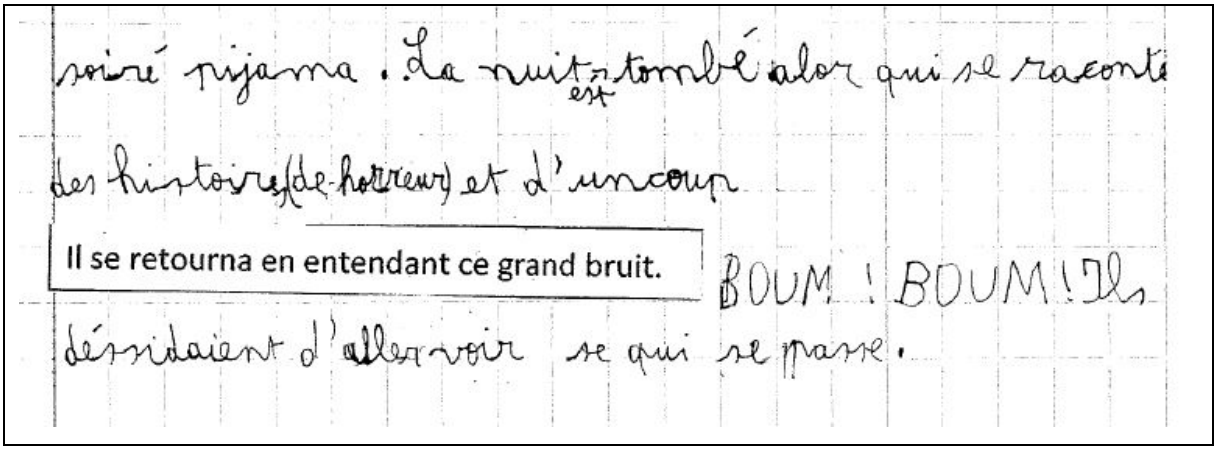
L'étude des formes de la première mention du référent du SN démonstratif *ce grand bruit* montre enfin que quelques textes comportent des onomatopées : 9% dans l'ensemble du corpus (soit 17 textes). On retrouve ce type de forme dans des proportions égales quel que soit le niveau de classe. L'emploi de ce type de forme se justifie par la nature même de l'onomatopée[55]. Dans notre corpus, il est intéressant de remarquer que les onomatopées apparaissent accompagnées du verbe *entendre* lorsque la première mention du référent est localisée avant P2, comme dans (95) :


Il étais une fois une petite fille qui s'appellait Rose
Rose qui habitait dans une joli maison blanche. Elle habitait dans
cette maison depuis longtemps. Son frère Eliot entendit
grrr. Il se retourna en entendant ce grand bruit. Un

Il était une fois une petite fille qui s'appelait Rose qui habitait dans une jolie maison blanche. *Elle habitait dans cette maison depuis longtemps.* Son frère Eliot entendit « Grrr ». *Il se retourna en entendant ce grand bruit.*

Figure 95 : Texte d'Olivia, CE2 U

Les onomatopées apparaissent seules lorsque la première mention est située après P2, tel que dans (96) avec *BOUM ! BOUM !* :


soirée pijama. La nuit est tombé alor qui se raconte
des histoires de horreur et d'un coup
Il se retourna en entendant ce grand bruit.
BOUM ! BOUM ! Ils
déssidaient d'aller voir se qui se passe.

[55] Selon la définition du Tlf, une onomatopée est : « un mot créé par imitation du son, évoquant l'être ou la chose que l'on veut nommer ».

La nuit est tombée alors qu'ils se racontaient des histoires (d'horreur) et d'un coup, *il se retourna en entendant ce grand bruit.* « BOUM ! BOUM ! » Ils décidèrent d'aller voir ce qu'il se passe.

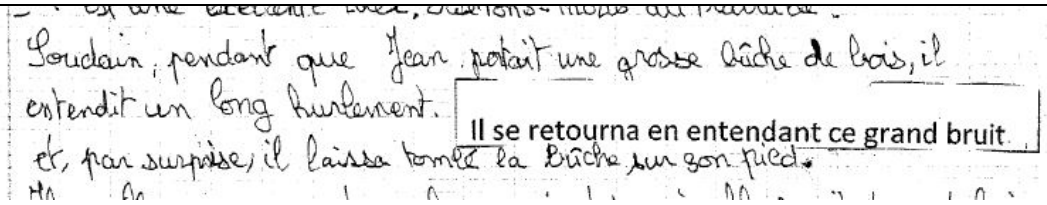
Figure 96 : Texte de Mathéo, CM1 T1

Étude des formes lexicales de M1 pour ce grand bruit

Lorsque les élèves emploient un SN pour introduire le référent, nous distinguons deux cas de figure :

- Le nom tête du SN employé en première mention, *bruit*, est le même que le nom tête du SN démonstratif *ce grand bruit*, comme dans l'exemple (91) *supra.*

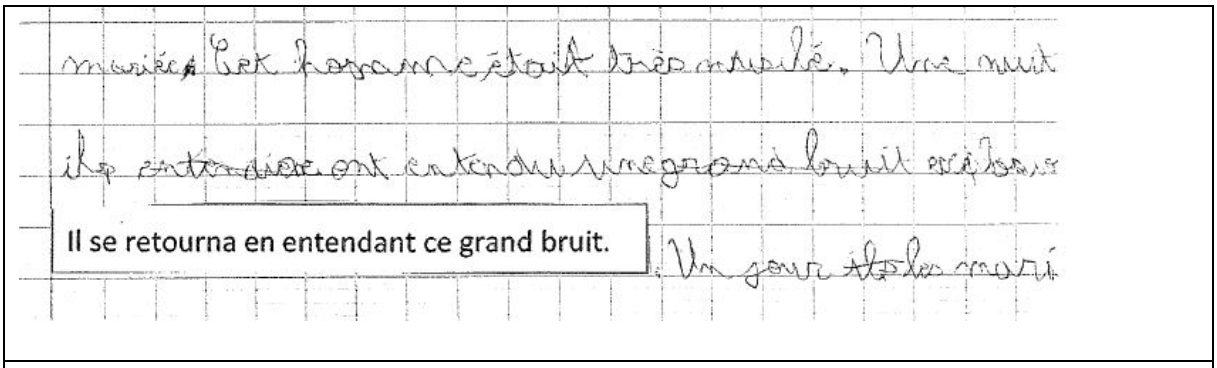
- Le nom tête du SN employé en première mention est différent du nom tête du SN démonstratif *ce grand bruit*. Il s'agit alors principalement d'un hyponyme du nom *bruit*, tel que *cri, grincement* ou *hurlement* comme dans (97) :



Soudain, pendant que Jean portait une grosse bûche de bois, il entendit un long hurlement. *Il se retourna en entendant ce grand bruit.* Et, par surprise, il laissa tomber la bûche sur son pied.

Figure 97 : Texte de Lucile, CM1 T1

On trouve également des noms contenus dans le champ notionnel du nom *bruit*, comme par exemple *une explosion* dans (98). Ce type de SN apparaît alors avec le verbe *entendre* et le SN ainsi choisi répond à la fois à la contrainte sémantique portant sur le SN *bruit* mais aussi à celle portant sur le SN expansé *grand bruit*. En effet, une explosion produit non seulement du bruit, mais c'est un bruit que l'on peut qualifier de *grand*.



Cet homme était très musclé. Une nuit ils ont entendu une explosion. *Il se retourna en entendant ce grand bruit.*

Figure 98 : Texte de Mathis, CM1 T1

On remarque d'autre part dans ce texte que le SN indéfini *une explosion* est un ajout en cours d'écriture. La première version de la phrase contenait le même nom tête que dans P2 (*grand bruit*) mais avec un article indéfini. Le remplacement du SN *un grand bruit* par le SN *une explosion* montre la prise de conscience du fait qu'il est préférable de ne pas reprendre les mêmes termes pour référer à une même entité (cf. méta-règle de répétition de Charolles, 1978 ; chapitre 1, section 1.1.). La révision du texte (Hayes & Flower, 1980 ; cf. chapitre 3, section 3.1.4.) témoigne du fait que la gestion de ce type d'élément n'est pas naturelle chez l'élève en cours d'apprentissage.

11.1.2. Etude des expansions de la première mention du référent du SN démonstratif *ce grand bruit*

Comme vu précédemment, dans P2, le SN démonstratif à introduire, *ce grand bruit*, est expansé à l'aide du syntagme adjectival *grand*. Cet ajout implique qu'il ne s'agit pas de n'importe quel type de bruit. Cet élément doit être pris en compte pour assurer la cohésion sémantique du texte. À cette fin, une forme dont le nom tête est différent de *bruit* peut être employée, mais il faut qu'elle comporte en elle-même la notion de grand bruit, comme par exemple *explosion* (cf. *supra,* figure 98). Cependant, pour référer au SN démonstratif *ce grand bruit*, un SN expansé à l'aide d'un modifieur peut également être employé. Il faut alors que le modifieur donne au mot *bruit* l'importance attendue.

L'étude des formes de la première mention du référent du SN démonstratif *ce grand bruit* montre que si, dans leur majorité, les textes du corpus ne présentent pas d'expansion,

ceux qui en présentent une seule (10%) sont plus nombreux que ceux qui en présentent deux (4 textes), comme le montre le tableau 34.

Nombre d'expansions de M1 pour *ce grand bruit*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
1 expansion	10	16%	4	7%	5	8%	19	**10%**
2 expansions	1	2%	0	0%	3	5%	**4**	2%
pas d'expansion	53	83%	52	93%	57	88%	162	88%
Total	64	100%	56	100%	65	100%	185	100%

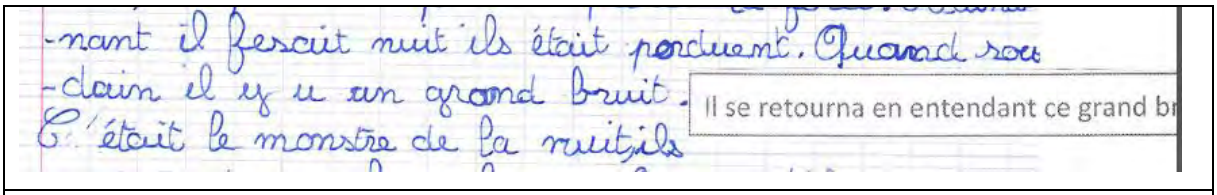
Tableau 34 : Nombre d'expansions de la première mention du référent du SN démonstratif *ce grand bruit*

Les types d'expansions utilisés peuvent être de différentes sortes : syntagme adjectival, proposition relative, syntagme prépositionnel, comme on le voit dans le tableau 35.

Type de première expansion de M1 pour *ce grand bruit*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
syntagme adjectival	10	91%	3	75%	7	88%	20	**87%**
proposition relative	1	9%	0	0%	0	0%	**1**	4%
syntagme prépositionnel	0	0%	1	25%	1	13%	**2**	9%
Total	11	100%	4	100%	8	100%	23	100%

Tableau 35 : Type de première expansion dans les formes de la première mention du référent du SN démonstratif *ce grand bruit*

L'étude des types d'expansion de la première mention du référent du SN démonstratif *ce grand bruit* montre que lorsque le SN introduisant la première mention est expansé, il l'est le plus fréquemment à l'aide d'un syntagme adjectival : 87% sur l'ensemble des premières expansions. Le syntagme adjectival le plus rencontré est *grand,* comme on le voit dans (99) :



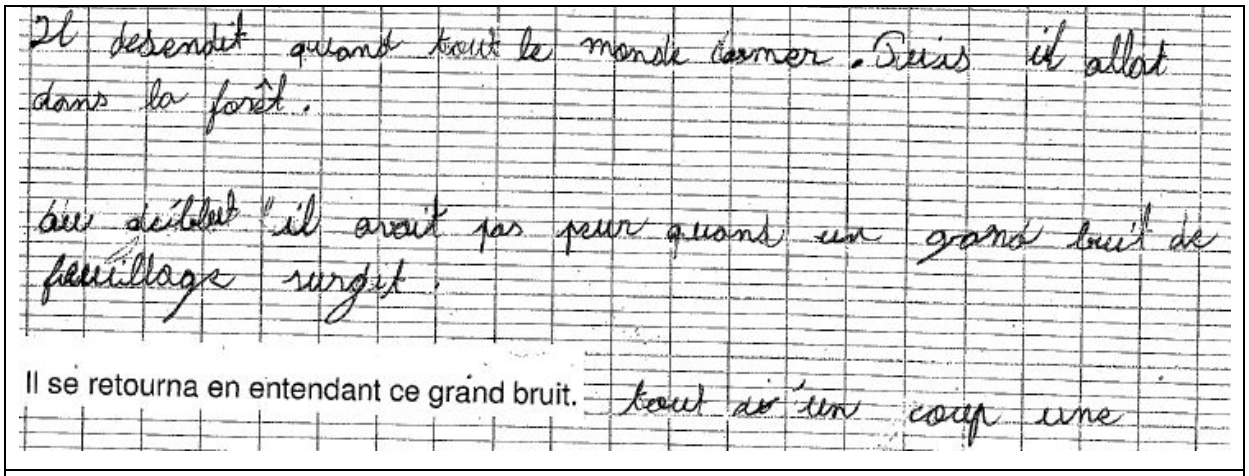
Quand soudain il y eut un grand bruit. *Il se retourna en entendant ce grand bruit.*

Figure 99 : Texte de Juliette, CE2 T

Parmi les autres types d'expansion, on trouve également :

- des syntagmes prépositionnels (2 textes), tels que *un bruit de claquement de porte* (Aïnhoa, CM1 T1) ou *le grincement des escaliers* (Maëva, CM2 B)
- une proposition relative : *ce bruit qui ressemblait à celui de l'histoire* (Adrien, CE2 T)

D'autre part, dans les formes qui comportent deux expansions (4 textes), on retrouve deux propositions relatives et un syntagme prépositionnel, comme dans (100) avec le SN expansé *un grand bruit de feuillage* :



Il descendit quand tout le monde dormait. Puis il alla dans la forêt. Au début, il n'avait pas peur, quand un grand bruit de feuillage surgit. *Il se retourna en entendant ce grand bruit.*

Figure 100 : Texte de Nathan, CM2 S

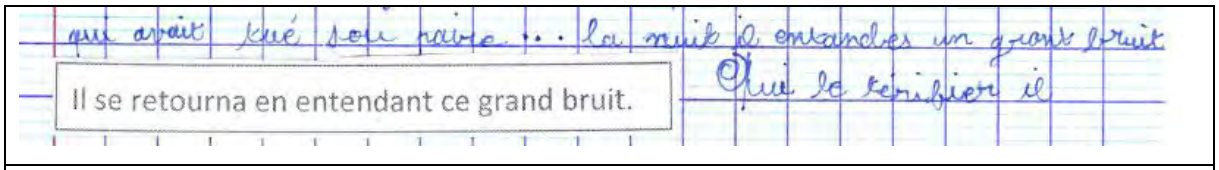
11.1.3. Etude des fonctions syntaxiques de la première mention du référent du SN démonstratif *ce grand bruit.*

Le SN démonstratif *ce grand bruit* dans P2 est en position de complément de verbe. De la même manière que pour le SN démonstratif *cette maison* dans P1 (cf. chapitre 10), le SN démonstratif *ce grand bruit* est une occurrence qui doit être prise en compte dans les textes car le démonstratif appelle une introduction préalable du référent. Mais dans P2, le pronom personnel *il* nécessite également l'introduction d'un référent. Il y a donc dans la même phrase deux occurrences qui nécessitent une introduction préalable : un référent humain animé (le personnage masculin) et un référent non humain inanimé (le grand bruit). Comme le caractère « humain » ou « animé » d'une entité augmente sa saillance, par rapport à une entité « non animée » ou un animal (Lyons, 1980 ; cf. chapitre 1, section 1.2.4.), l'élève devrait avoir tendance à garder la même « hiérarchie » que celle de P2 (stratégie de non changement de rôle, cf. chapitre 3, section 3.1.) et le référent du SN démonstratif *ce grand bruit* devrait alors se retrouver le plus fréquemment en position de complément de verbe, plus rarement en position de sujet syntaxique.

Cette hypothèse est confirmée par nos résultats dans une certaine mesure. En effet, 28% des élèves du corpus placent la première mention du référent en position de complément de verbe, comme l'illustre le tableau 36 et l'exemple (101) ci-après.

Fonctions syntaxiques de M1 pour *ce grand bruit*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
complément de verbe	23	36%	14	25%	15	23%	52	**28%**
sujet	6	9%	1	2%	4	6%	**11**	6%
autre	20	31%	12	21%	20	31%	52	**28%**
phrase nominale	2	3%	4	7%	6	9%	**12**	6%
non pertinent	13	20%	25	45%	20	31%	58	31%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 36 : Différents types de fonctions syntaxiques prises par la première mention du référent du SN démonstratif *ce grand bruit*



La nuit il entendait un grand bruit. *Il se retourna en entendant ce grand bruit.*

Figure 101 : Texte d'Eliott 2, CE2 T

Dans cet exemple, les référents du pronom personnel *il* et du SN démonstratif *ce grand bruit* sont dans les mêmes positions syntaxiques que dans P2 : le pronom personnel *il* est en position de sujet syntaxique et le SN indéfini *un grand bruit* en position de complément de verbe.

Les textes dans lesquels la première mention du référent est placée en position de sujet syntaxique sont minoritaires : 11 textes dans l'ensemble du corpus (cf. exemple (100) *supra).*

Cependant, la catégorie *autre* regroupe également 28% des textes du corpus. Il s'agit des phrases entières qui expliquent la provenance du bruit (cf. ce chapitre, section 11.1.1.).

Enfin, les phrases nominales représentent 12 textes du corpus. Cette catégorie regroupe tous les cas où la forme de la première mention est une onomatopée seule (cf. ce chapitre, section 11.1.1.).

11.2. Maintien du référent du SN démonstratif *ce grand bruit*

Après avoir introduit le référent, l'élève peut choisir de le maintenir tout au long de son discours. Nous allons donc étudier à présent comment les élèves maintiennent la référence au SN démonstratif *ce grand bruit* dans leur texte. En 11.2.1., nous dressons une vue rapide et générale du maintien de la référence à travers l'étude du nombre de mentions des CDR et de leur longueur moyenne En 11.2.2., nous étudions de manière plus précise différents cas de figure, selon qu'il s'agit d'anaphore (1 mention + P1), de CDR minimales (2 mentions + P1) ou de CDR plus longues (3 mentions ou plus + P1).

11.2.1. Vue d'ensemble des CDR incluant le SN démonstratif *ce grand bruit*

Le SN démonstratif *ce grand bruit* représente plutôt une occurrence secondaire dans l'histoire, notamment par son caractère non humain et non animé, en opposition, nous l'avons déjà vu, avec les personnages humains des phrases imposées. En outre, l'emploi du passé simple dans P2 avec *il se retourna* implique que le bruit entendu est une forme de complication ponctuelle dans le récit mais qui devrait être vite résolue. De ce fait, on s'attend à ce que les CDR incluant le SN démonstratif *ce grand bruit* soient plutôt courtes.

Nos résultats confirment cette hypothèse. La longueur moyenne des CDR incluant le SN démonstratif *ce grand bruit* augmente globalement en fonction du niveau scolaire, mais le nombre moyen de maillons par CDR est d'environ 2 :

- 1,77 maillons au CE2
- 1,65 maillons au CM1
- 1,99 maillons au CM2

De manière générale, l'étude du nombre de mentions (ou maillons) de la CDR incluant le SN démonstratif *ce grand bruit* montre que les CDR comportant une seule mention en plus de P1 (soit deux mentions au total) sont les plus fréquentes : 52% dans l'ensemble du corpus, comme le montre le tableau 37.

Nombre de mentions de la CDR de *ce grand bruit*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
1 mention + P2	42	66%	25	45%	29	45%	96	**52%**
2 mentions + P2	7	11%	4	7%	11	17%	22	**12%**
3 mentions + P2	2	3%	2	4%	4	6%	**8**	4%
4 mentions + P2	0	0%	1	2%	1	2%	**2**	1%
9 mentions + P2	0	0%	0	0%	1	2%	1	1%
P2 = seule mention	13	20%	24	43%	19	29%	56	**30%**
Total	64	100%	56	100%	65	100%	185	100%

Tableau 37 : Nombre de mentions dans la CDR incluant le SN démonstratif *ce grand bruit*

On constate également que les textes dans lesquels le SN démonstratif *ce grand bruit* de P2 est la seule mention du référent représentent 30% des textes du corpus. Ce fort taux peut s'expliquer par la multiplicité des contraintes qui s'appliquent au SN démonstratif *ce grand bruit.* En effet, non seulement il représente une entité secondaire pour l'histoire, mais de surcroît, le sémantisme du nom *bruit* et les contraintes qui lui sont associées à travers l'expansion *grand* font de cette occurrence celle qui est sans doute la plus difficile à gérer pour les élèves.

Après les deux cas que nous venons de citer, les CDR les plus fréquentes sont celles présentant 2 mentions en plus de celle de P2 (soit 3 mentions au total) : 12% du corpus. Enfin, les CDR plus longues sont minoritaires : les CDR avec 3 mentions + P2 représentent 8 textes et les CDR avec 4 mentions + P2 représentent 2 textes.

Nous remarquons un cas particulier qui se trouve au CM2. Chaque type de CDR est représenté jusqu'à 4 mentions + P2. Toutefois, au-delà de cette longueur, il n'y a aucun cas si ce n'est une CDR de 9 mentions + P2, soit 10 mentions au total.

11.2.2. Différentes configurations des CDR

Nous l'avons vu, notre corpus regroupe plusieurs configurations de CDR. Il peut s'agir d'anaphore ou de cataphore (1 mention + P2), de CDR minimales (2 mentions + P2) ou de CDR plus longues (3 mentions ou plus + P2). Chacune de ces configurations fait l'objet d'une étude plus précise.

Étude des cas d'anaphore ou de cataphore : 1 mention + P2

Le cas de figure le plus fréquent dans notre corpus lorsqu'il s'agit de maintenir la référence à l'occurrence *bruit* est le recours à la configuration suivante : 1 mention + P2 (cf. tableau *supra*). Ce type de configuration témoigne d'une gestion de la référence au SN démonstratif plutôt locale ou « en arrière-plan ». En effet, le petit nombre de mentions du référent ne permet pas d'en faire une entité véritablement présente dans le discours ou bien de construire une histoire entière dont elle serait le sujet thématique. Toutefois, comme nous

l'avons déjà vu, le sémantisme du SN démonstratif *ce grand bruit* et sa nature non humaine non animée n'en font pas une occurrence principale pour l'histoire.

Dans la configuration « 1 mention + P2 », on trouve deux cas de figure :

- des anaphores (Reichler-Béguelin, 1988 ; cf. chapitre 2) : des textes dans lesquels il y a une mention du référent avant P2 et celle déjà présente dans P2, ou *anaphore*

- des cataphores (cf. chapitre 2, section 2.1.) : des textes dans lesquels le SN démonstratif *ce grand bruit* de P2 est la première mention du référent. La source du référent est alors localisée après P2.

Dans chacun de ces cas, la forme de la mention du référent (hors celle de P2) n'est pas la même selon qu'elle se trouve avant P2 (anaphore) ou après P2 (cataphore).

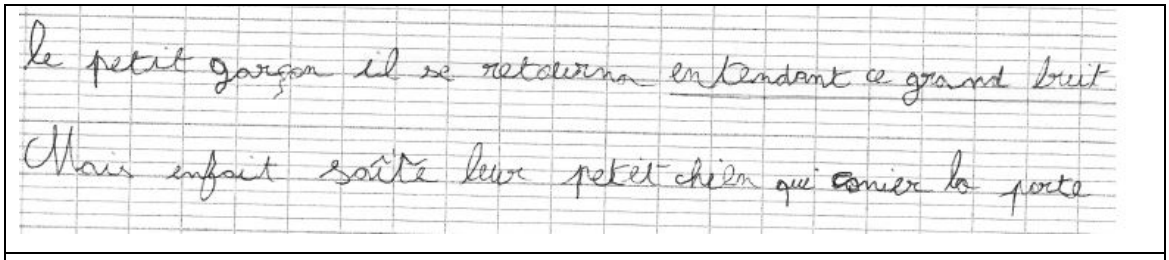
En effet, dans les cas d'anaphore, la mention du référent située avant P2 est la première, elle sert donc d'introduction. Les caractéristiques de la première mention du référent sont alors les mêmes que celles analysées dans la section 11.1.1, lorsque nous avons étudié l'introduction du référent : la forme linguistique la plus fréquente est le SN indéfini en position de complément de verbe (pour une illustration par des exemples de textes d'élève, nous renvoyons aux exemples déjà vus en section 11.1.1.).

Dans les cas de cataphore, la première mention du référent est le SN démonstratif *ce grand bruit* contenu dans P2. La source du référent est alors localisée après P2. Dans ces cas, la forme la plus fréquente de la source du référent est une phrase entière qui explique la provenance du bruit et/ou sa nature : 64% des cas de cataphore, comme l'indique le tableau 38.

Formes de M1 de *ce grand bruit* dans les cas de cataphore								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN indéfini	0	0%	2	20%	1	14%	3	7%
SN défini	2	8%	0	0%	0	0%	2	5%
SN démonstratif	1	4%	2	20%	0	0%	3	7%
SN pronominal	1	4%	0	0%	0	0%	1	2%
onomatopée	2	8%	4	40%	0	0%	6	**14%**
phrase explicative	19	76%	2	20%	6	86%	27	**64%**
Total	25	100%	10	100%	7	100%	42	100%

Tableau 38 : Formes de la première mention du référent du SN démonstratif *ce grand bruit* dans les cas de cataphore (P2 + 1 mention après P2)

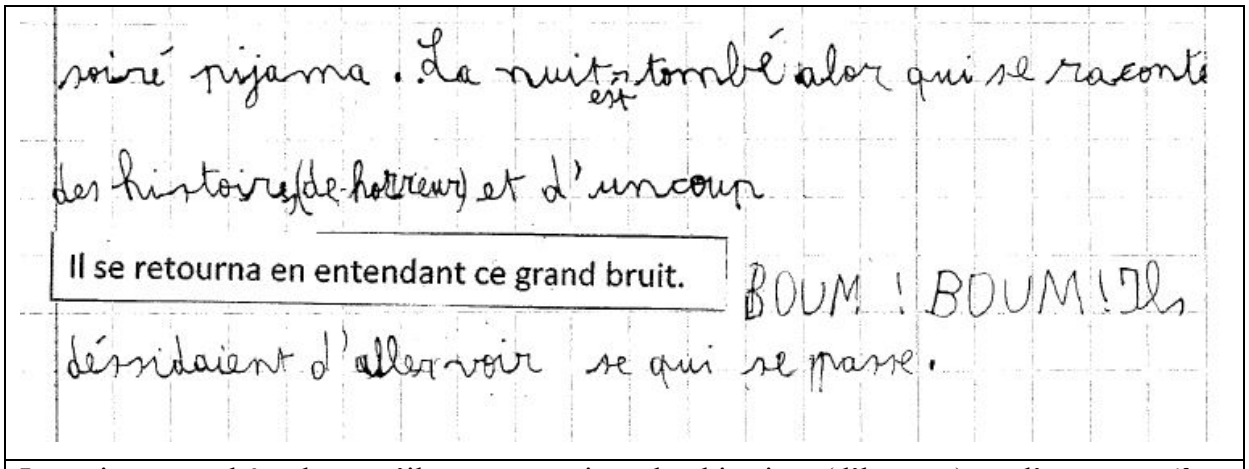
Par exemple dans (102), aucune mention du référent du SN démonstratif *ce grand bruit* n'apparaît avant P2, mais la phrase qui suit explique que le bruit entendu provient d'un chien qui heurte une porte :


le petit garçon il se retourna en tendant ce grand bruit
Mais enfait sait leur petit chien qui cognier la porte

le petit garçon *il se retourna en entendant ce grand bruit.* Mais en fait c'était leur petit chien qui cognait la porte.

Figure 102 : Texte de Rose, CE2 U

Les onomatopées sont la deuxième forme la plus fréquente de la source du SN démonstratif *ce grand bruit*: 14% des cas de cataphore, comme l'illustre (103) avec *BOUM ! BOUM !* .


soiré pijama. La nuit est tombé alor qui se raconte des histoires (de horreur) et d'uncoup

Il se retourna en entendant ce grand bruit.

BOUM ! BOUM ! Ils décidaient d'aller voir se qui se passe.

La nuit est tombée alors qu'ils se racontaient des histoires (d'horreur) et d'un coup, *il se retourna en entendant ce grand bruit.* « BOUM ! BOUM ! ». Ils décidèrent d'aller voir ce qu'il se passe.

Figure 103 : Texte de Mathéo, CM1 T1

Dans les cas de cataphore, lorsque l'on met à part les formes de phrases entières, la fonction syntaxique la plus fréquente est le complément de verbe, à l'image de la position syntaxique occupée par le SN démonstratif *ce grand bruit* dans P2.

Étude des CDR minimales : 2 mentions + P2

Le deuxième cas de figure le plus fréquent dans notre corpus lorsqu'il s'agit de maintenir la référence à l'occurrence *grand bruit* est le recours à la configuration suivante : 2 mentions + P2 (cf. tableau 37 *supra*), c'est-à-dire 3 mentions au total. Il s'agit là de la configuration minimale dans laquelle on peut parler de CDR (Schnedecker, 1997 ; cf. chapitre 2, section 2.2.). On parle donc de CDR minimale, en opposition avec les CDR contenant plus de 3 maillons que nous aborderons plus tard.

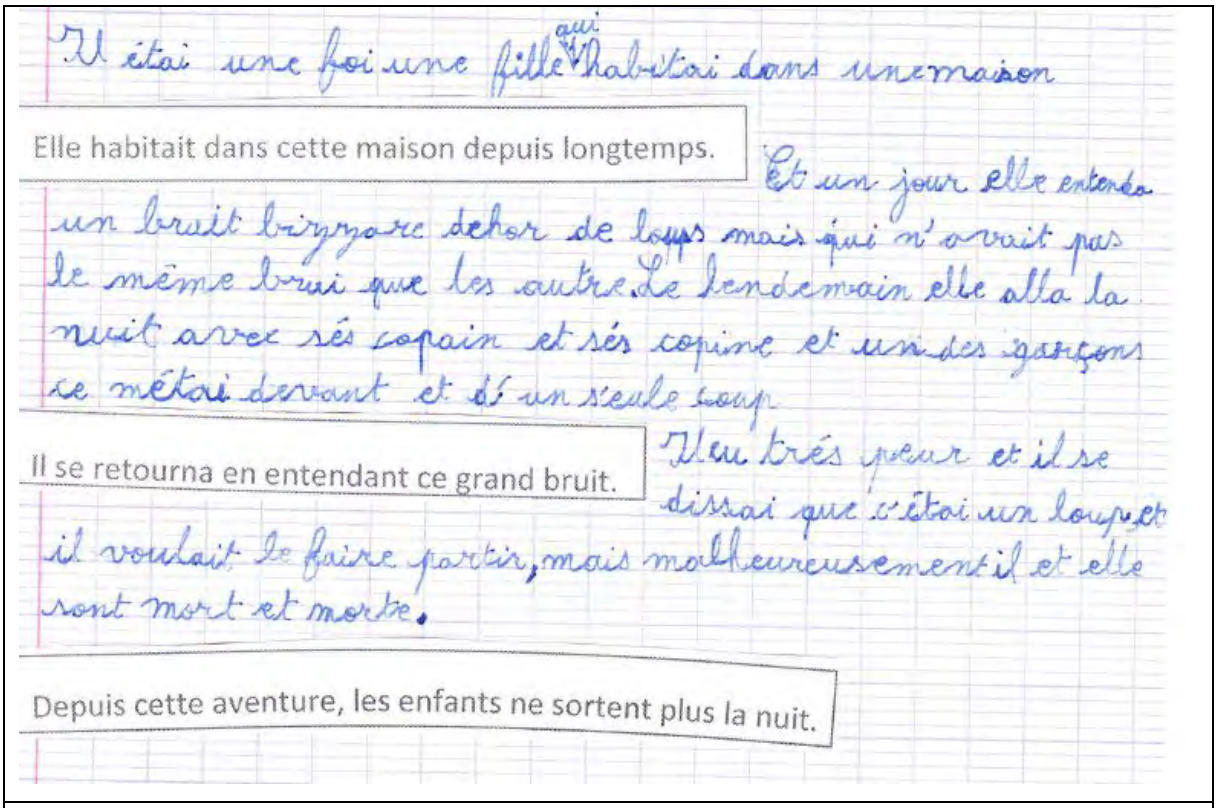
Les CDR témoignent d'une gestion de la référence plus globale. En effet, en mentionnant le référent du SN démonstratif *ce grand bruit* plusieurs fois en dehors de P2, l'élève commence à construire son « personnage » (en l'occurrence l'entité *grand bruit*) en lui octroyant une place plus importante. Encore une fois, le caractère non humain et non animé de l'occurrence *grand bruit* ne lui permet pas *a priori* d'avoir la même importance que les personnages humains de la consigne. Néanmoins, étant mentionnée plusieurs fois,

l'occurrence *grand bruit* devient un élément secondaire qui garde une relative importance, comme un élément déclencheur d'une ou plusieurs actions.

Les travaux qui se sont intéressés au maintien de la référence ont montré que le SN défini a toute sa place en tant que deuxième maillon dans une CDR (Landragin & *al.*, 2015 ; cf. chapitre 2, section 2.2.4.) car c'est une forme de moyenne accessibilité (Ariel, 1980 ; cf. chapitre 1, section 1.2.2.) qui peut être employée lorsque le référent a déjà été introduit au préalable, ce qui est le cas ici. À la lecture du tableau 39, on constate pourtant que le SN défini n'est que la deuxième forme la plus fréquente dans les cas de CDR minimales et représente 23% des textes, comme l'illustre l'exemple (104) avec *le bruit*.

Formes de M2 de *ce grand bruit* dans les CDR minimales								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN défini	2	29%	2	50%	1	9%	5	**23%**
SN démonstratif	0	0%	0	0%	3	27%	3	14%
phrase entière	3	43%	2	50%	7	64%	12	**55%**
onomatopée	2	29%	0	0%	0	0%	2	9%
Total	7	100%	4	100%	11	100%	22	100%

Tableau 39 : Formes de la deuxième mention du référent du SN démonstratif *ce grand bruit* dans les CDR minimales


Il étai une foi une fille qui habitai dans une maison

Elle habitait dans cette maison depuis longtemps.

Et un jour elle entenda un bruit bizyare dehor de loups mais qui n'avait pas le même brui que les autre. Le lendemain elle alla la nuit avec sés copain et sés copine et un des garçons se métai devant et d'un seule coup

Il se retourna en entendant ce grand bruit.

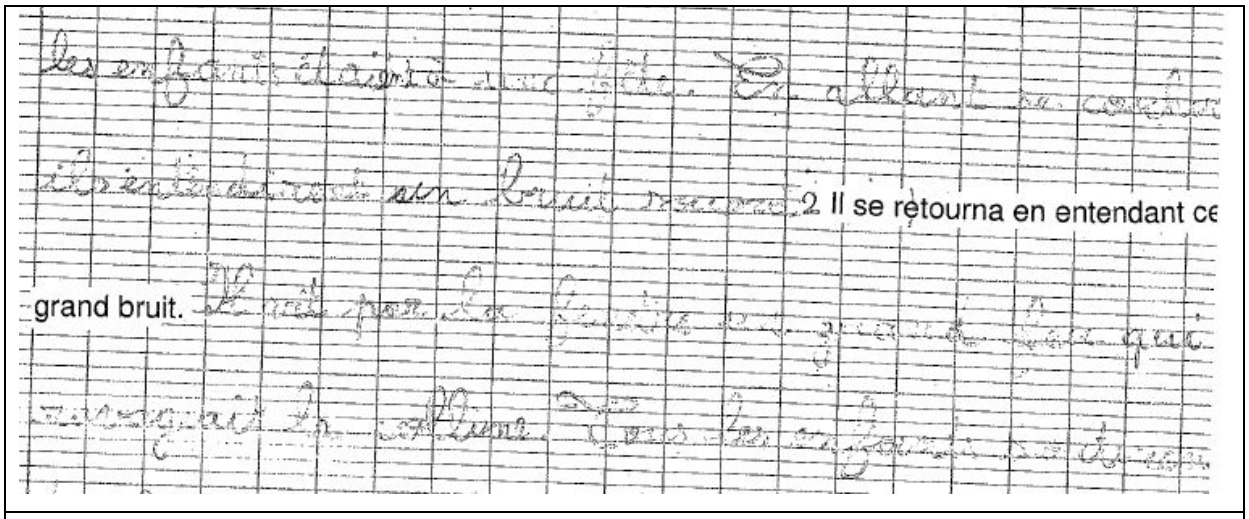
Ileu trés peur et il se dissai que c'étai un loup et il voulait le faire partir, mais malheureusement il et elle sont mort et morte.

Depuis cette aventure, les enfants ne sortent plus la nuit.

Il était une fois une fille qui habitait dans une maison. *Elle habitait dans cette maison depuis longtemps.* Et un jour elle entendit un bruit bizarre, dehors, de loup, mais qui n'avait pas le même bruit que les autres. Le lendemain elle alla la nuit avec ses copains et ses copines et un des garçons se (mit) devant et d'un seul coup, *il se retourna en entendant ce grand bruit.* Il eut très peur et il se disait que c'était un loup et il voulait le faire partir, mais malheureusement il et elle sont mort et morte. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 104 : Texte de Karl, CM1 T1

La forme de M2 la plus fréquente dans notre corpus est la phrase entière : 55% des cas de CDR minimales, comme on le voit dans (105) où le grand feu est la source du bruit entendu. Ce résultat peut s'expliquer par les particularités sémantiques attachées au SN démonstratif *ce grand bruit* (cf. ce chapitre, section 11.1.) qui incitent les élèves à expliquer la provenance et/ou la nature du bruit à l'aide d'une phrase.



En allant se coucher, ils entendirent un bruit sourd. *Il se retourna en entendant ce grand bruit.* Il vit par la fenêtre un grand feu qui ravageait la colline.

Figure 105 : Texte de Max, CM2 S

L'étude des fonctions syntaxiques de la deuxième mention du référent du SN démonstratif *ce grand bruit* dans les CDR minimales montre que, lorsqu'on excepte les phrases entières (regroupées sous la catégorie *autre* étant donné qu'elles n'ont pas de fonction syntaxique propre), on constate que la position de complément de verbe est la plus fréquente : 27% des cas de CDR minimales, comme on le voit dans le tableau 40.

Fonctions syntaxiques de M2 dans les CDR minimales								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
complément de verbe	2	29%	1	25%	3	27%	6	**27%**
sujet	1	14%	0	0%	1	9%	2	9%
autre	4	57%	3	75%	7	64%	14	64%
Total	7	100%	4	100%	11	100%	22	100%

Tableau 40 : Fonctions syntaxiques de la deuxième mention du référent du SN démonstratif *ce grand bruit* dans les CDR minimales

Comme nous l'avons vu dans l'étude des fonctions syntaxiques de la première mention du référent (cf. ce chapitre, section 11.1.3.), l'élève a tendance à garder la même

« hiérarchie » que celle de P2 et le référent du SN démonstratif se retrouve en position de complément de verbe, quel que soit le niveau de classe, plus rarement en position de sujet syntaxique.

Étude des CDR plus longues : 3 mentions (ou plus) + P2

Les textes présentant des CDR plus longues, c'est-à-dire 3 mentions (ou plus) + P2 sont moins nombreux (cf. tableau 36 *supra*) :

- 8 textes comportent 3 mentions + P2
- 2 textes comportent 4 mentions + P2
- 1 texte comporte 9 mentions + P2

Plus la CDR comporte de mentions du référent, moins les textes sont nombreux. Nous ne pourrions ici faire une revue détaillée de chacune des mentions du référent du SN démonstratif *ce grand bruit* comme nous l'avons fait jusqu'à présent. Nous nous contenterons donc d'une vue générale des compositions des CDR comportant 3 maillons (ou plus) + P1.

En ce qui concerne la forme des mentions du référent du SN démonstratif *ce grand bruit,* globalement, il n'y a pas de prédominance constatée d'une forme sur l'autre : les SN définis, démonstratifs, onomatopées et phrases explicatives sont aussi fréquents les uns que les autres. Toutefois, la comparaison par niveaux montre que les formes telles que les SN définis et démonstratifs ont plutôt tendance à apparaître chez les élèves les plus jeunes.

Les rares expansions constatées (2 textes) se trouvent exclusivement dans la troisième mention du référent. Il n'y en a pas au-delà. Toutes deux sont des syntagmes adjectivaux.

L'étude des fonctions syntaxiques des différentes mentions montre également qu'il n'y a pas de prédominance de l'une ou l'autre des fonctions : on retrouve indifféremment des formes en position de complément de verbe ou de sujet.

Nous retrouvons ces éléments dans le texte de Jules (CM2 B) que nous reproduisons dans son intégralité et qui comporte la plus longue CDR incluant le SN démonstratif *ce grand bruit* dans notre corpus.

Une petite fille rentrant de l'école. Entendis des gens disent

qu'une maison entier n'était pas loin ~~de tout~~ de l'école.

La maison entier était au numéro 13 rue du Stylo.

Ça lui disait quelque chose parce que sa

1 Elle habitait dans cette maison depuis longtemps.

En rentrant [illegible]

avait un

elle ~~et~~ ~~[illegible]~~ ~~[illegible]~~ frère qui avait 6 ans. Elle alla dans

sa chambre faire sa [illegible] devoirs. Quand elle [illegible] [illegible]

fini, descendi les escaliers, elle entendi un bruit. Puis [illegible] [illegible]

Elle

Puis

Puis un autre bruit, bubababobom ! Sont frère crier

a sa sœur pour savoir qu'est-ce que c'était.

2 Il se retourna en entendant ce grand bruit.

Ce bom !!!! Puis ils

sortir de la maison mais sa venait enfaite de

dehors. Ils courut dans la rue et vu des ~~casserol~~

casserole

volé, c'était sa les bom bom !!! ~~qu'ils~~ C'était des

fantômes qui faisait sa !

3 Depuis cette aventure, les enfants ne sortent plus la nuit.

parce qu'il

ont peur des fantômes.

Une petite fille rentrant de l'école entendit des gens dire qu'une maison hantée n'était pas loin de l'école. La maison hantée était au numéro 13 rue du Stylo. Ca lui disait quelque chose parce qu'*elle habitait dans cette maison depuis longtemps.* En rentrant chez elle, (elle) avait un frère qui avait 6 ans. Elle alla dans sa chambre faire ses devoirs. Quand elle eut fini, elle descendit les escaliers puis elle entendit un bruit. « Babom ! ». Puis un autre bruit, « bababababom ! ». Son frère criait à sa sœur pour savoir ce que c'était. *Il se retourna en entendant ce grand bruit.* « Bom !!!! ». Puis ils sortirent de la maison mais ça venait en fait de dehors. Ils coururent dans la rue et virent des casseroles voler, c'était ça les « bom bom » !!! C'était des fantômes qui faisaient ça ! *Depuis cette aventure, les enfants ne sortent plus la nuit,* parce qu'ils ont peur des fantômes.

Figure 106 : Texte de Jules, CM2 B

Dans ce texte, on peut considérer que le référent du SN démonstratif *ce grand bruit* est mentionné plusieurs fois avant P2, sous la forme du SN indéfini *un bruit* pour la première mention, conformément à ce qui est attendu, et sous la forme d'onomatopées, *babom,* ou du pronom *ça* pour les mentions suivantes. Dans les mentions du référent qui apparaissent après P2, on retrouve les mêmes formes que précédemment. En outre, on voit également une phrase entière qui explique la provenance du bruit : *[…] et virent des casseroles voler, c'était ça les « bom, bom ».* Le grand nombre de maillons de la CDR, en gardant le référent dans le focus d'attention du lecteur, donne au texte sa *texture* (Halliday & Hasan, 1976 ; cf. chapitre 1, section 1.1.) et montre une gestion globale de la référence.

11.3. Bilan de l'étude du référent du SN démonstratif *ce grand bruit*

De manière générale, l'étude des formes de la première mention du référent du SN démonstratif *ce grand bruit* montre que :

- la première mention du référent apparaît principalement sous la forme d'une phrase entière qui explique la provenance du bruit. Ce type d'introduction est moins fréquent lorsque le niveau de classe augmente. Le SN indéfini est la deuxième forme la plus fréquente, quel que soit le niveau de classe.
- la première mention du référent apparaît le plus fréquemment en position de complément de verbe, quel que soit le niveau de classe.

- les formes lexicales employées en première mention peuvent être le même nom tête que celui du SN démonstratif *ce grand bruit*, ce qui est le plus fréquemment constaté quel que soit le niveau de classe, ou bien dépendre du sémantisme du nom *bruit* et se réaliser soit sous la forme d'un hyponyme, soit sous la forme d'un autre nom dans le champ notionnel de *bruit*. Ces deux derniers types de forme sont plus représentés lorsque le niveau de classe augmente.
- lorsque la forme de la première mention est expansée, elle l'est majoritairement à l'aide d'une seule expansion. On constate que les formes expansées sont plus fréquentes au CE2 que dans les autres niveaux de classe. Le type d'expansion utilisé est principalement un syntagme adjectival, tous niveaux confondus.

L'étude du maintien de la référence au SN démonstratif *ce grand bruit* montre que dans l'ensemble du corpus, la longueur moyenne des CDR incluant le SN démonstratif *ce grand bruit* est de 2 maillons, c'est-à-dire 1 mention + celle de P2. Dans cette configuration, les cas d'anaphore (la source du référent est placée avant P2) présentent les mêmes caractéristiques que celles étudiées dans le cas de l'introduction du référent : mis à part les textes où le référent est une phrase entière expliquant la provenance du bruit, cas les plus fréquents, la forme qui apparaît le plus est le SN indéfini en position de complément de verbe. Dans les cas de cataphore (la source du référent est placée après P2), la forme la plus fréquente est la phrase explicative, suivi des onomatopées. Les formes telles que les SN définis, démonstratifs et pronoms sont minoritaires mais apparaissent le plus souvent en position de complément de verbe.

Dans les CDR minimales, c'est-à-dire comportant 2 mentions + celle de P3 (soit 3 maillons), la forme de la deuxième mention est majoritairement une phrase explicative. Les SN définis sont toutefois la deuxième forme la plus fréquente et on les retrouve principalement en position de complément de verbe.

Enfin, dans les CDR plus longues, qui restent tout de même minoritaires dans le corpus (quelques textes), il n'y a pas de réelle prédominance d'une forme sur l'autre ou d'une fonction sur l'autre. La plupart des cas de figure sont représentés.

CHAPITRE 12 : Enchâssement des référents de la troisième phrase imposée : *Depuis cette aventure, les enfants ne sortent plus la nuit* dans l'ensemble du texte

Dans la tâche-problème que doivent réaliser les élèves de notre étude, la troisième et dernière phrase à insérer (P3) est la suivante :

Depuis cette aventure, les enfants ne sortent plus la nuit.

Par sa position et sa composition, cette phrase a un statut particulier. D'une part, l'anaphore résomptive *depuis cette aventure* et le SN défini *la nuit* suggèrent que les enfants ont vécu un ou plusieurs évènements qualifiables d'*aventure*, source de danger et/ou de peur, qui les conduit à ne plus sortir la nuit. P3 semble donc contenir un caractère de moralité générale, accentué par l'emploi du verbe au présent. D'autre part, les élèves doivent décider des référents du SN défini *les enfants* : soit il s'agit d'une reprise des personnages de P1 et P2, mentionnés sous la forme des pronoms personnels *elle* et *il*, soit il s'agit d'un emploi plus « générique » de l'article défini, dans le sens où le SN *les enfants* réfère à toute une catégorie d'enfants, par exemple tous les enfants d'une aire géographique. Les élèves peuvent également intégrer à leur récit des personnages d'enfants qui ne sont pas imposés par la consigne et les regrouper sous le SN défini *les enfants*.

En outre, le SN défini *la nuit* implique que les évènements racontés se passent pendant la nuit. Cette indication temporelle doit être prise en compte dans le récit. Ainsi, la phrase finale P3 contient un caractère conclusif qui lui permet d'être employée comme dernière phrase du texte produit. C'est une phrase *fermoir* (Marandin, 1986) : elle induit la construction d'un récit mettant en scène des enfants, apparaissant comme une *aventure* et conduisant à l'abandon des sorties nocturnes, probablement du fait d'une scène effrayante (Garcia-Debanc, 2012). Tous ces éléments nécessitent donc une planification à rebours du texte de l'élève (Hayes & Flower, 1980 ; Schneuwly, 1988 ; cf. chapitre 3, section 3.1.4) car ils obligent à réfléchir aux différentes étapes de sa construction.

Aussi, les différents éléments mentionnés feront chacun l'objet d'une section de ce chapitre. En 12.1., nous allons donc étudier le traitement du référent du SN défini *les enfants*, en analysant à la fois l'introduction du référent et son maintien tout au long du discours. En

12.2., nous examinerons comment l'indication temporelle *la nuit* est prise en compte, à savoir s'il s'agit d'une gestion globale ou plutôt locale du référent. Enfin, en 12.3., nous analyserons de manière plus qualitative l'enchâssement de P3 dans l'ensemble du texte des élèves, et notamment la réalisation de l'anaphore résomptive *depuis cette aventure*.

12.1. Le SN défini *les enfants*

Dans la troisième phrase imposée de la consigne (P3), *Depuis cette aventure, les enfants ne sortent plus la nuit*, le SN défini *les enfants* nécessite l'introduction préalable d'un référent. En 12.1.1., nous nous attachons donc à décrire les diverses formes de la première mention du référent, ses éventuelles expansions ainsi que les fonctions syntaxiques qu'il peut occuper. En 12.1.2., nous étudions comment les élèves maintiennent le référent tout au long du discours, en examinant notamment la nature des CDR employées.

12.1.1. Introduction du référent du SN défini *les enfants*

De manière générale, dans notre corpus, on observe trois localisations possibles de l'expression référentielle qui introduit le référent (ou *source*), comme le montre le tableau 41 :

- La source (ou référent) est introduite avant la phrase imposée et la référence est maintenue ou non après
- La source n'est pas introduite avant la phrase imposée mais elle est localisée après.
- La source est absente : la seule mention du référent est celle de la phrase imposée

P3 étant la dernière phrase à insérer dans le récit, sa valeur de phrase *fermoir* (Marandin, 1986) implique que le référent soit préférentiellement introduit avant. Étant donné le caractère conclusif de P3, on ne s'attend pas à ce que le texte se poursuive après, ni à ce que la source de la première mention du référent soit localisée après P3.

Localisation de la source de *les enfants*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
avant P3 et maintien après	41	64%	38	68%	36	55%	115	**62%**
après P3	2	3%	0	0%	1	2%	**3**	2%
source absente	21	33%	18	32%	28	43%	67	**36%**
Total	64	100%	56	100%	65	100%	185	100%

Tableau 41 : Localisation de la source du SN défini *les enfants*

Pourtant, si les résultats de notre étude montrent que l'introduction du référent avant P3 est le cas le plus fréquent (62% des textes de notre corpus), on constate de manière anecdotique qu'il y a 3 textes dans notre corpus dans lesquels la source du SN défini *les* enfants est localisée après P3. Il s'agit de cataphores (cf. chapitre 2, section 2.1.1.1.) : le SN défini *les enfants* de P3 est la première mention du référent, qui est repris par la suite. Cette manière de localiser le référent semble plus fréquente lorsque les élèves sont plus jeunes : 2 textes au CE2, 1 texte au CM2. Notons qu'il n'y en a aucun au CM1.

De surcroît, sur l'ensemble du corpus, 36% des textes n'attribuent aucune source au référent du SN défini *les enfants*. P3 est la seule mention de l'occurrence. Toutefois, cette catégorie nécessite que l'on s'y attarde un instant. En effet, le caractère de moralité générale de P3 et les spécificités du SN *les enfants* nous conduisent à distinguer trois cas :

- P3 est la seule mention du référent mais l'on comprend par inférence à la lecture du texte dans son ensemble que le SN défini *les enfants* a pour référent des personnages de l'histoire. C'est le cas par exemple dans (107) où le SN défini *les enfants* réfère aux personnages de Fleur, Rouge et Vulcain bien que dans le texte, ils ne soient réunis à aucun moment sous la forme d'un groupe d'individus :

Il était une fois trois fantastique. Don Fleur une elfe, elle
rivaient dans une petite maison. Elle habitait dans cette
maison depuis longtemps. Fleur avait 12 ans. Il y avai
aussi Rouge! Fille du diable, elle n'aimait pas son père
qui passait ses journée à haïr les mort. Rouge à 13 ans.

Les deux amie se baladaient, lorsque un garçon apparu. Il
se retourna en entendant ce grand bruit. Cracc! Il ne
savait guère d'où venait se bruit... Aïïe! Il poussa
quelques branche avent d'apperservoir Fleur urlente de
douleur. La nuit començait à tombé.

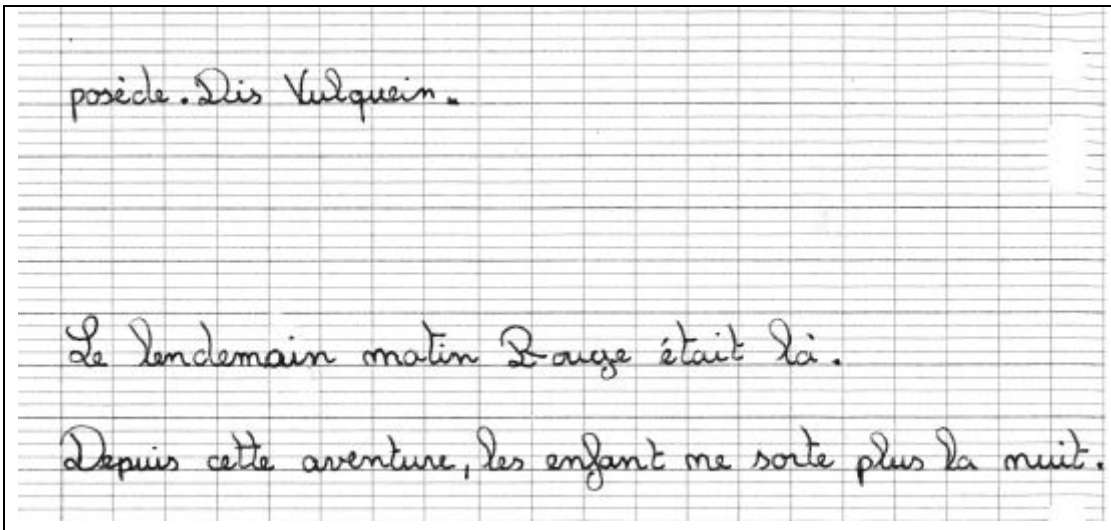
-(Fleur) Comment t'appel tu?

-(Le garçon) Je m'appel Vulquein.

-(Vulquein) Tu est toute seul.

(Fleur) Non on à kidnaper mon amie!

Rentront avent que les fantôme nous


posède. Dis Vulquein.

Le lendemain matin Rouge était là.

Depuis cette aventure, les enfant ne sorte plus la nuit.

Il était une fois trois fantastiques. Dont Fleur une elfe, elle vivait dans une petite maison. *Elle habitait dans cette maison depuis longtemps.* Fleur avait 12 ans. Il y avait aussi Rouge ! Fille du Diable, elle n'aimait pas son père qui passait ses journées à haïr les morts. Rouge a 13 ans. Les deux amies se baladaient lorsqu'un garçon apparu. *Il se retourna en entendant ce grand bruit.* « Crac ! ». Il ne savait guère d'où venait ce bruit… « Aïe ! ». Il poussa quelques branches avant d'apercevroi Fleur hurlante de douleur. La nuit commençait à tomber.

(Fleur) : Comment t'appelles-tu ?

(Le garçon) : Je m'appelle Vulcain.

(Vulcain) : Tu es toute seule ?

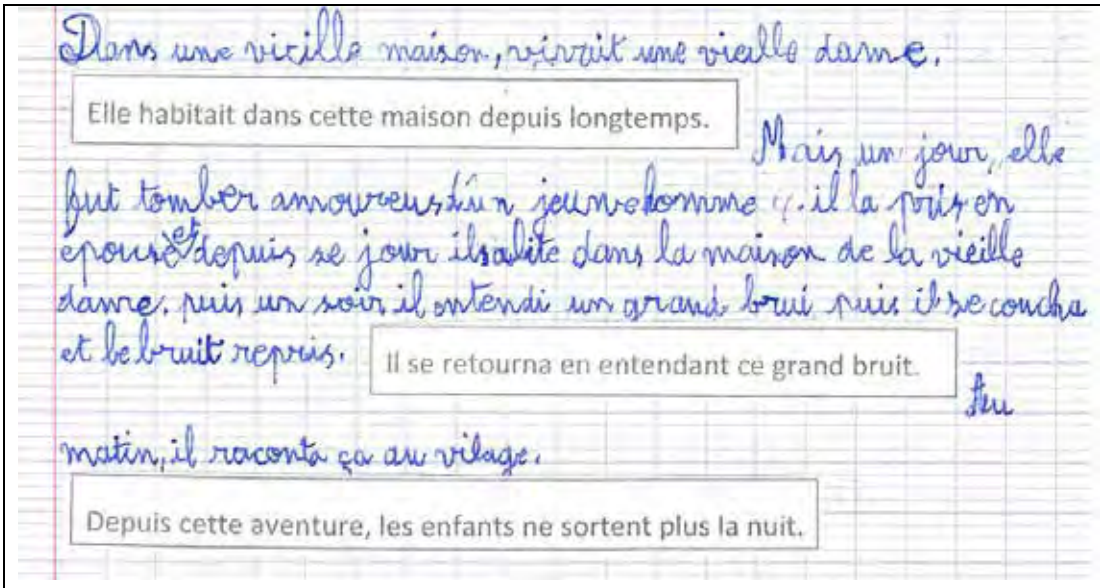
(Fleur) : Non, on a kidnappé mon amie !

- Rentrons avant que les fantômes nous possèdent, dit Vulcain.

Le lendemain matin, Rouge était là. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 107 : texte de Cindy, CM1 T1

- P3 est la seule mention du référent mais l'on comprend par inférence à la lecture du texte dans son ensemble que le SN défini *les enfants* réfère à toute une catégorie d'enfants, tels que dans (108) où la mention du village avant P3 permet d'inférer que le référent du SN défini *les enfants* est l'ensemble des enfants du village :


Dans une vieille maison, vivait une vieille dame.

Elle habitait dans cette maison depuis longtemps.

Mais un jour, elle fut tomber amoureuse d'un jeune homme et il la prit en épouse et depuis se jour ils habite dans la maison de la vieille dame. puis un soir, il entendi un grand bruit puis il se coucha et le bruit reprit.

Il se retourna en entendant ce grand bruit.

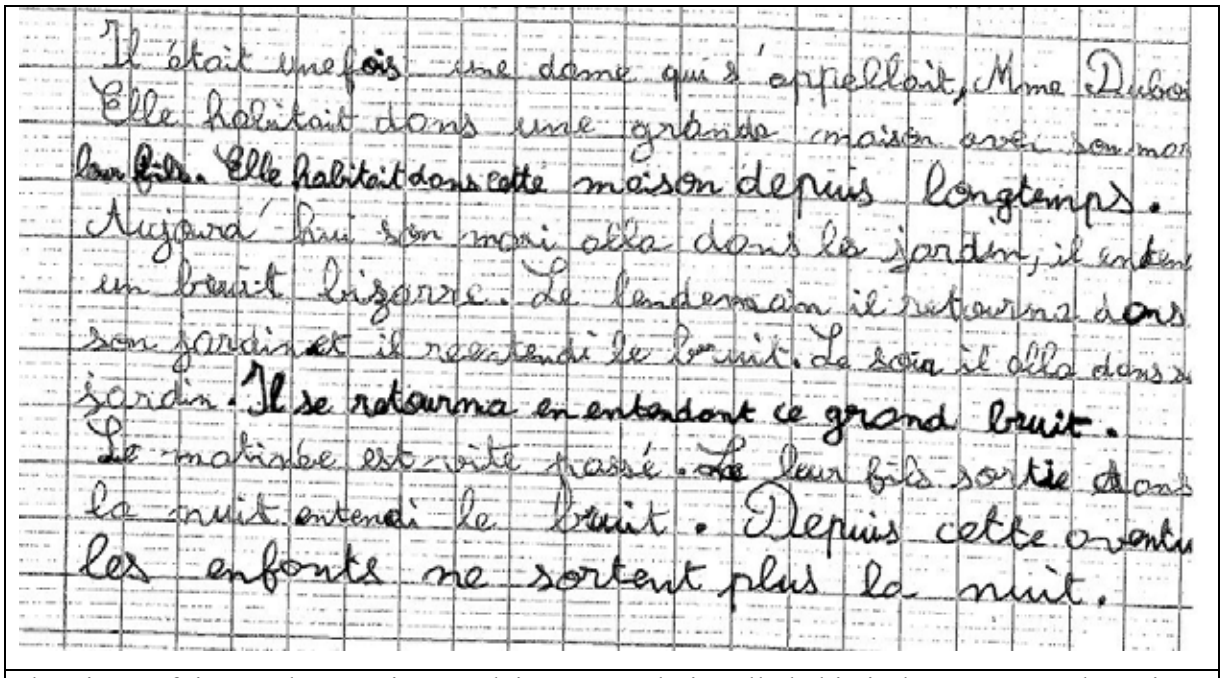
Au matin, il raconta ça au vilage.

Depuis cette aventure, les enfants ne sortent plus la nuit.

Dans une vieille maison, vivait une vieille dame. *Elle habitait dans cette maison depuis longtemps.* Mais un jour, elle tomba amoureuse d'un jeune homme. Il la prit pour épouse et depuis ce jour, ils habitent dans la maison de la vieille dame. Puis un soir, il entendit un grand bruit. Il se coucha et le bruit reprit. *Il se retourna en entendant ce grand bruit.* Au matin, il raconta ça au village. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 108 : Texte de Jolin, CE2 T

- P3 est la seule mention du référent et l'on ne peut lui trouver aucune source. C'est le cas par exemple dans (109) où aucun référent ne peut être assimilé au SN défini *les enfants*, même par inférence :



Il était une fois une dame qui s'appellait Mme Dubois. Elle habitait dans une grande maison avec son mari et leur fils. *Elle habitait dans cette maison depuis longtemps.* Aujourd'hui, son mari alla dans le jardin, il entendit un bruit bizarre. Le lendemain il retourna dans son jardin et il réentendi le bruit. Le soir il alla dans son jardin. *Il se retourna en entendant ce grand bruit.* La matinée est vite passé. Leur fils sortie dans la nuit entendi le bruit. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 109 : Texte de Lisa, CM2 T

La fréquence de ces trois cas est détaillée dans le tableau 42 ci-dessous. Lorsque P3 est la seule mention du référent dans le texte, on constate que 61% des textes ne contiennent aucune source du référent, même par inférence. On constate également que les textes dans lesquels il n'y a aucune mention du référent du SN défini *les enfants* mais dans lesquels on peut inférer qu'il s'agit d'une catégorie d'enfants sont plus fréquents (24%) que ceux dans lesquels on peut inférer qu'il s'agit de personnages de l'histoire (15%).

Types de référents pour P3 = seule mention de *les enfants*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
P3 = personnages	3	14%	4	22%	3	11%	10	**15**%
P3 = générique	6	29%	3	17%	7	25%	16	**24**%
P3 = aucun référent	12	57%	11	61%	18	64%	41	**61**%
Total	21	100%	18	100%	28	100%	67	100%

Tableau 42 : Types de référents quand P3 est la seule mention du SN défini *les enfants*

Après avoir dressé une vue d'ensemble des différentes manières de gérer la référence, nous allons examiner plus en détail les formes de la première mention du référent du SN défini *les enfants*.

12.1.1.1. Étude des formes de la première mention du référent du SN défini *les enfants*

Les travaux qui se sont intéressés à l'introduction des référents chez les enfants (notamment Hickmann, 1980, 1984, 1987 ; cf. chapitre 3, section 3.1.2.1.), et les adultes (notamment Corblin, 1985 ; cf. chapitre 2, section 2.2.) ont montré que la manière la plus ordinaire d'introduire un nouveau référent dans un discours est d'utiliser un SN indéfini. Or dans notre corpus, du fait des particularités de la tâche-problème proposée (cf. ce chapitre, introduction et section 12.1.1.), la première mention du référent du SN défini *les enfants* peut être de différentes natures :

- Le groupe nominal *les enfants* est constitué par les personnages de P1 et P2, c'est-à-dire les référents des pronoms personnels *elle* et *il*, qui sont réunis à un moment donné dans le texte sous la forme du SN défini *les enfants* ou du pronom personnel *ils*
- Le groupe nominal *les enfants* est constitué par des personnages différents de ceux de P1 et P2, intégrés au texte par l'élève sous la forme de SN indéfinis, numéraux ou possessifs
- Le groupe nominal *les enfants* réfère à tous les enfants d'une catégorie. Dans ce cas, si le référent est introduit avant P3, il peut se trouver sous la forme d'un SN défini, le défini ayant ici une valeur générique

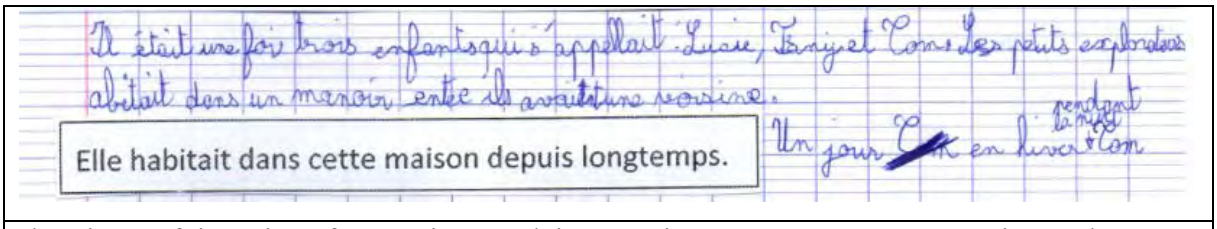
Le tableau 43 reproduit les différents types de formes de la première mention, quelle que soit la nature de l'occurrence *les enfants*.

Formes de M1 pour *les enfants*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN indéfini	1	2%	5	9%	3	5%	9	5%
SN numéral	15	23%	17	30%	12	18%	44	**24%**
SN possessif	4	6%	6	11%	8	12%	18	10%
SN défini	8	13%	5	9%	9	14%	22	**12%**
pronom	15	24%	5	9%	5	8%	25	**14%**
P3 = seule mention	21	33%	18	32%	28	43%	67	36%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 43 : Formes de la première mention du référent du SN défini *les enfants*

Les différentes formes d'introduction du référent du SN défini les enfants

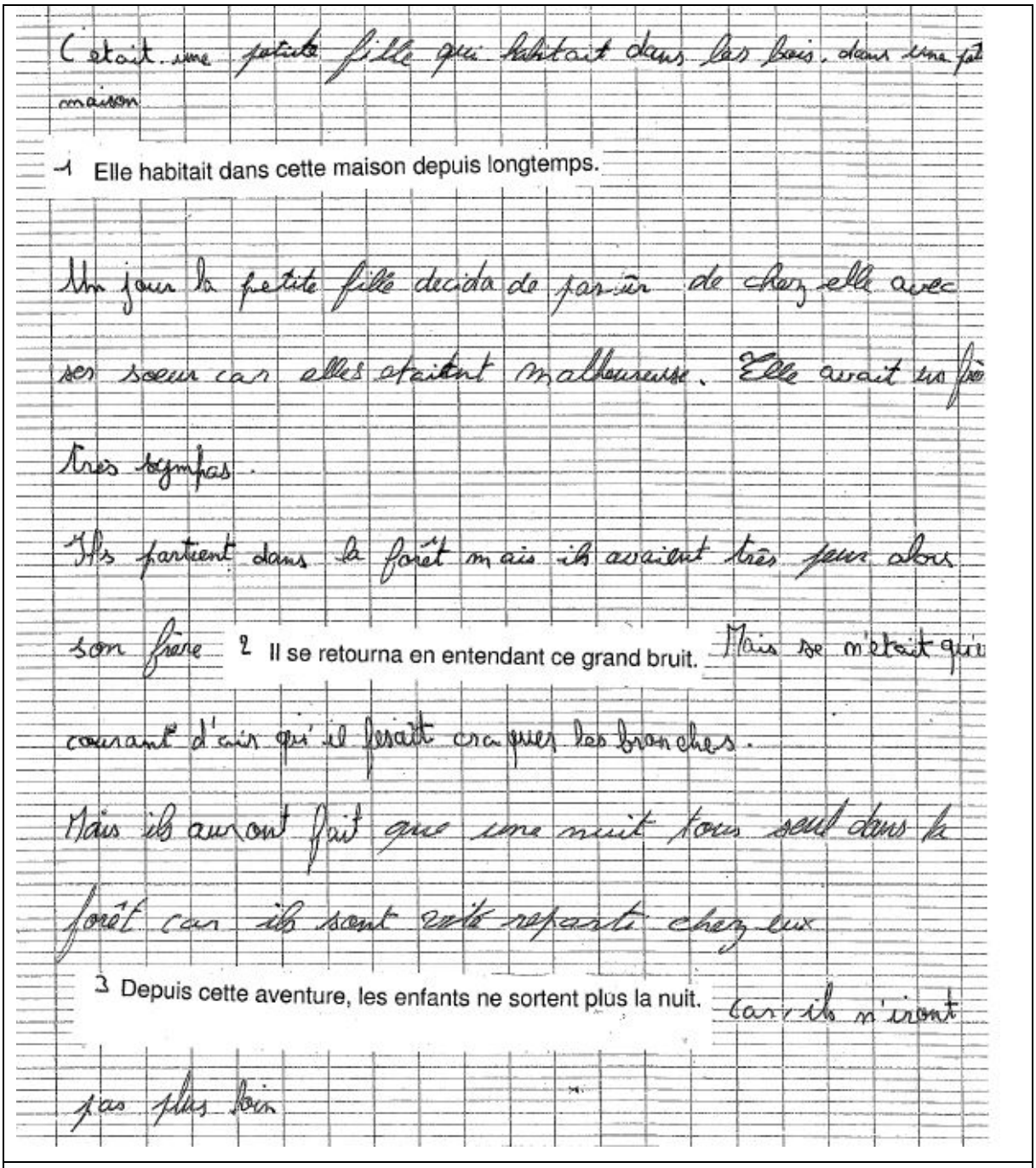
À la lecture du tableau, mis à part les cas où P3 est la seule mention du référent, on constate que le SN numéral est la forme de la première mention la plus fréquente : 24% des textes du corpus, comme l'illustre l'exemple (110) avec le SN numéral expansé *trois enfants qui s'appelaient Lucie, Fanny et Tom* :



Il était une fois trois enfants qui s'appelaient Lucie, Fanny et Tom. Les petits explorateurs habitaient dans un manoir hanté, ils avaient une voisine. *Elle habitait dans cette maison depuis longtemps.*

Figure 110 : Texte de Nina, CE2 L

La deuxième forme la plus fréquente est le pronom : 14% dans l'ensemble du corpus. Il s'agit de textes dans lesquels, une fois les référents des pronoms personnels *elle* et *il* introduits, on recourt au pronom pluriel *ils* pour les réunir et constituer un groupe qui devient le référent du SN défini *les enfants*, comme dans (111).



C'était une petite fille qui habitait dans les bois, dans une petite maison.

1 Elle habitait dans cette maison depuis longtemps.

Un jour la petite fille decida de partir de chez elle avec ses soeur car elles etaitent malheureuse. Elle avait un frère très sympas.

Ils partient dans la forêt mais ils avaient très peur alors son frère 2 Il se retourna en entendant ce grand bruit. Mais se n'était qu'un courant d'air qui il faisait craquer les branches.

Mais ils auront fait que une nuit tous seul dans la forêt car ils sont vite repartis chez eux

3 Depuis cette aventure, les enfants ne sortent plus la nuit. Car ils n'iront pas plus loin

C'était une petite fille qui habitait dans les bois, dans une petite maison. *Elle habitait dans cette maison depuis longtemps.* Un jour, la petite fille décida de partir de chez elle avec ses sœurs car elles étaient malheureuses. Elle avait un frère très sympa. Ils partent dans la forêt mais ils avaient très peur alors son frère, *il se retourna en entendant ce grand bruit.* Mais ce n'était qu'un courant d'air qui faisait craquer les branches. Mais ils auront fait qu'une nuit tout seuls dans la forêt car ils sont vite repartis chez eux. *Depuis cette aventure, les enfants ne sortent plus la nuit.* Car ils n'iront pas plus loin.

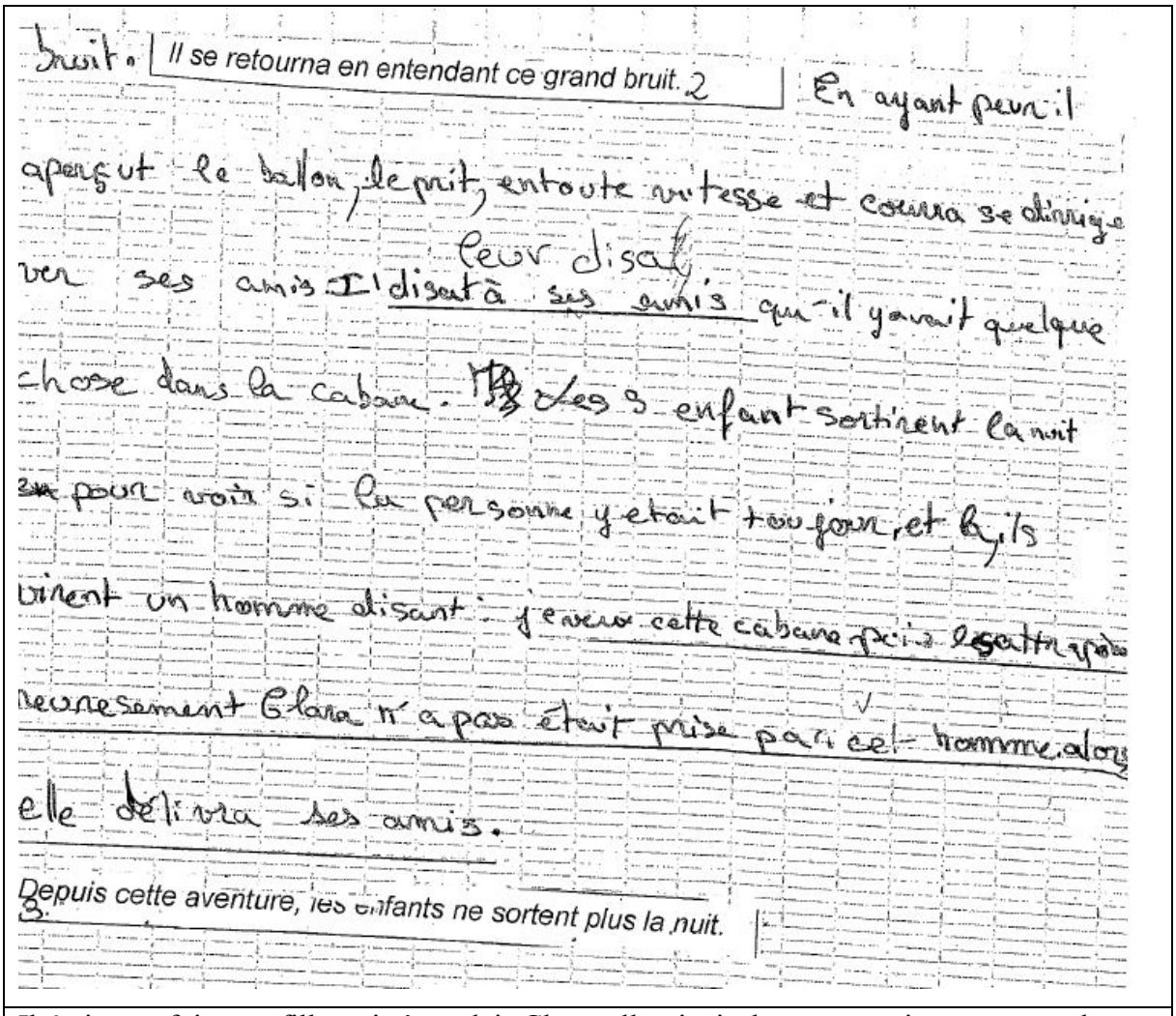
Figure 111 : Texte sans nom 1, CM2 S

Le SN défini est la troisième forme de la première mention du référent du SN défini *les enfants* la plus fréquente : 12% des textes du corpus. Comme nous l'avons vu, nous pouvons distinguer les cas où le SN défini est un regroupement de plusieurs personnages identifiés, comme dans (112) où le SN défini *les enfants* réfère au groupe constitué par Clara et les autres enfants :

Il était une fois une fille qui s'appellait Clara, elle vivait dans une maison toute seule les parents étaient mort noyés.

Elle habitait dans cette maison depuis longtemps.

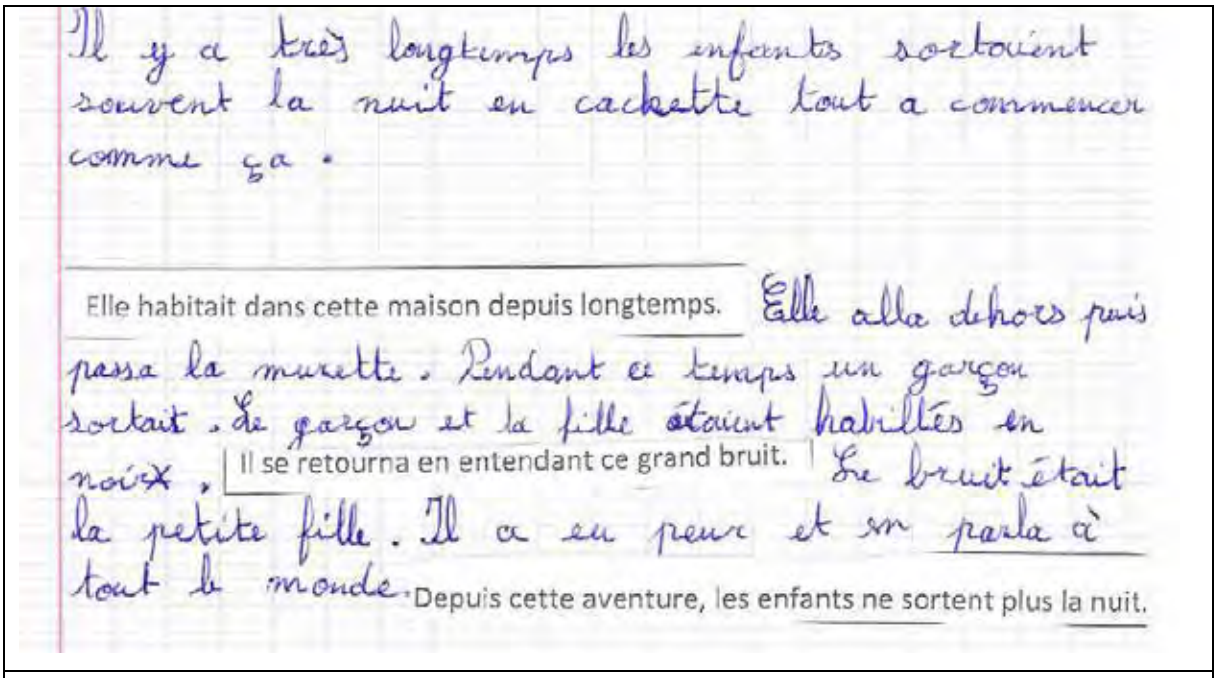
Un jour en sortant de chez elle pour cueillir des champignons et de la menthe, dans la forêt elle vit 1 fille et 3 garçons qui lui dirent: veux tu venir avec nous jouer, on veut devenir tes amis ? Et si on jouer à se faire des passes avec un ballon? J'en est un dans notre cabane. Le jeune garçon alla alors, retourner dans cette cabane. cinq minute plus tard, quand il rentrait dans la cabane il entendait un


bruit. Il se retourna en entendant ce grand bruit. 2 En ayant peur il aperçut le ballon, le prit, en toute vitesse et courra se diriger ver ses amis. Il leur dit qu'il y avait quelque chose dans la cabane. Les 3 enfant sortirent la nuit pour voir si la personne y était toujour, et là, ils virent un homme disant : je veux cette cabane puis l'attrapèrent heureusement Clara n'a pas était prise par cet homme alors elle délivra ses amis.
3. Depuis cette aventure, les enfants ne sortent plus la nuit.

Il était une fois une fille qui s'appelait Clara, elle vivait dans une maison toute seule, ses parents étaient morts noyés. *Elle habitait dans cette maison depuis longtemps.* Un jour en sortant de chez elle pour cueillir des champignons et de la menthe, dans la forêt elle vit 1 fille et 3 garçons qui lui dirent : « veux-tu venir avec nous jouer, on veut devenir tes amis ? Et si on jouait à se faire des passes avec un ballon ? J'en ai un dans notre cabane ». Le jeune garçon alla alors retourner dans cette cabane. Cinq minutes plus tard, quand il rentra dans la cabane, il entendit un bruit. *Il se retourna en entendant ce grand bruit*. En ayant peur il aperçut le ballon, le prit, en toute vitesse et courut se diriger vers ses amis. Il leur dit qu'il y avait quelque chose dans la cabane. Les 3 enfants sortirent la nuit pour voir si la personne y était toujours, et là, ils virent un homme disant : « Je veux cette cabane. » puis l'attrapèrent. Heureusement, Clara n'a pas été prise par cet homme alors elle délivra ses amis. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 112 : Texte d'Inaya, CM2 B

et les cas où le SN défini est employé pour sa valeur générique, comme dans l'exemple (113) avec le SN défini *les enfants* de la première phrase :


Il y a très longtemps les enfants sortaient souvent la nuit en cachette tout a commencer comme ça.

Elle habitait dans cette maison depuis longtemps. Elle alla dehors puis passa la murette. Pendant ce temps un garçon sortait. Le garçon et la fille étaient habillés en noir. Il se retourna en entendant ce grand bruit. Le bruit était la petite fille. Il a eu peur et en parla à tout le monde. Depuis cette aventure, les enfants ne sortent plus la nuit.

Il y a très longtemps, les enfants sortaient souvent la nuit en cachette, tout a commencé comme ça.
Elle habitait dans cette maison depuis longtemps. Elle alla dehors puis passa la murette. Pendant ce temps, un garçon sortait. Le garçon et la fille étaient habillés en noir. *Il se retourna en entendant ce grand bruit.* Le bruit était la petite fille. Il a eu peur et en parla à tout le monde. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 113 : Texte d'Amélie, CE2 P

12.1.1.2. Etude des expansions de la première mention du référent du SN défini *les enfants*

Mis à part les cas où le SN défini *les enfants* réfère à toute une catégorie d'individus, l'emploi de l'article défini dans le SN marque la valeur spécifique de l'occurrence : on sait qu'il s'agit de tel groupe d'enfants et pas d'un autre. On ne s'attend donc pas à ce que les mentions du référent soient expansées. Cependant, quelques textes présentent un ou plusieurs modifieurs qui sont autant d'ajouts d'informations à propos de l'entité ainsi dénotée (Charolles, 2002).

L'étude des formes de la première mention du référent du SN défini *les enfants* montre que si, dans leur grande majorité, les textes du corpus ne présentent pas d'expansion, ceux qui en présentent une seule (12 textes) sont plus nombreux que ceux qui en présentent deux (1 texte, au CM2).

Les types d'expansions utilisés peuvent être de différentes sortes : syntagme adjectival, proposition relative, syntagme prépositionnel ou nom propre. Tous les types d'expansions sont représentés dans des proportions égales (23% pour chaque type), excepté la proposition relative qui est légèrement plus fréquente (1 texte de différence). Dans le texte qui comporte deux expansions, la deuxième expansion est également une proposition relative. Dans les types de première expansion, on retrouve des formes telles que :

- Pour les syntagmes adjectivaux : *deux*, *grands et courageux*, *orphelins*
- Pour les propositions relatives : *qui s'appelaient…* (x2), *qui allaient demander des bonbons, qui se promenaient dans la forêt*
- Pour les syntagmes prépositionnels : *et un frère, de la région, du quartier,*

12.1.1.3. Etude des fonctions syntaxiques de la première mention du référent du SN défini *les enfants*

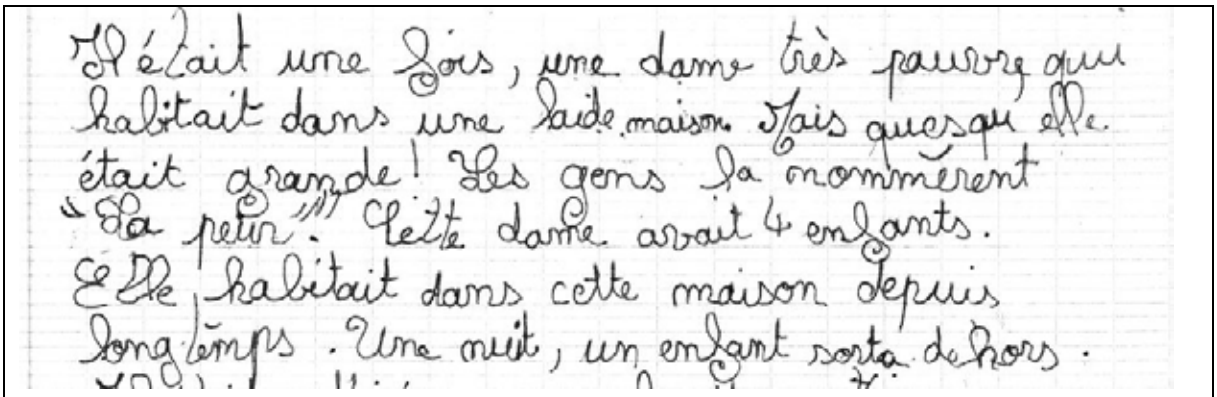
Dans leur texte, les élèves sont amenés à introduire le référent du SN défini *les enfants*. Nous l'avons vu, lorsque la première mention d'un référent apparaît, c'est une information totalement nouvelle (cf. *statut de l'information*, chapitre 1, section 1.2.) qui représente le thème du propos (cf. *progression thématique*, chapitre 1, section 1.2.3.). Dans ce cas, il est naturel que cette nouvelle information se retrouve en position initiale dans l'énoncé et prenne la place de sujet syntaxique, d'autant plus que le caractère humain et animé de l'occurrence *les enfants* augmente sa saillance (Lyons, 1980 ; Pattabhiraman, 1993 ; cf. chapitre 1, section 1.2.4.). Cependant, étant donné les différentes natures de l'occurrence *les enfants* (catégorie d'individus ou regroupement de personnages), d'autres positions syntaxiques peuvent être envisagées.

Cette hypothèse est confirmée par nos résultats, comme le montre le tableau 44 : 31% des élèves emploient la première mention du référent du SN défini *les enfants* en position de

sujet syntaxique (cf. (113) *supra*) et 24% en position de complément de verbe comme dans (114) (reproduit partiellement seulement).

Fonctions syntaxiques de M1 pour *les enfants*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
sujet	22	34%	19	34%	16	25%	57	**31%**
complément de verbe	16	25%	11	20%	17	26%	44	**24%**
construction présentative	2	3%	2	4%	2	3%	**6**	3%
complément du nom	2	3%	4	7%	1	2%	**7**	4%
phrase nominale	1	2%	2	4%	1	2%	**4**	2%
P3 = seule mention	21	33%	18	32%	28	43%	67	36%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 44 : Différents types de fonctions syntaxiques prises par la première mention du référent du SN défini *les enfants*


Il était une fois, une dame très pauvre qui habitait dans une laide maison. Mais quesque elle était grande! Les gens la nommèrent "La peur". Cette dame avait 4 enfants. Elle habitait dans cette maison depuis longtemps. Une nuit, un enfant sorta dehors.

Il était une fois, une dame très pauvre qui habitait dans une laide maison. Mais qu'est-ce qu'elle était grande ! Les gens la nommèrent « La peur ». Cette dame avait 4 enfants. *Elle habitait dans cette maison depuis longtemps.* Une nuit, un enfant sortit dehors.

Figure 114 : Texte de Lana, CM2 M

De manière plus marginale, on trouve également des premières mentions dans des constructions présentatives (cf. (110) *supra*) : 6 textes dans le corpus, ainsi qu'en position de complément du nom (7 textes) ou dans des phrases nominales (4 textes).

12.1.2. Maintien du référent du SN défini *les enfants*

Dans les cas où le référent du SN défini *les enfants* est introduit dans le texte avant P3, il peut être maintenu tout au long du discours. Nous allons donc étudier à présent comment les élèves maintiennent la référence au SN défini *les enfants* dans leur texte. En 12.1.2.1., nous dressons une vue générale du maintien de la référence à travers l'étude du nombre de mentions des CDR et de leur longueur moyenne. En 12.1.2.2., nous étudions de manière plus précise différents cas de figure, selon qu'il s'agit d'anaphore (1 mention + P1), de CDR courtes (2 mentions + P1) ou de CDR plus longues (3 mentions ou plus + P1).

12.1.2.1. Vue d'ensemble des CDR incluant le SN défini *les enfants*

De manière générale, l'étude du nombre de mentions (ou maillons) de la CDR incluant le SN défini *les enfants* montre que la CDR la plus longue comporte 17 maillons + celui de P3, soit 18 maillons au total, comme l'illustre le tableau 45. On constate par ailleurs que les CDR les plus longues (7 mentions et au-delà) sont présentes dans les textes de CM1 et CM2 mais pas dans ceux de CE2.

Nombre de mentions de la CDR de *les enfants*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
P3=seule mention	21	33%	18	32%	28	43%	67	36%
1 mention + P3	15	**23%**	5	9%	7	**11%**	27	**15%**
2 mentions + P3	7	11%	3	5%	5	8%	15	8%
3 mentions + P3	9	14%	8	**14%**	7	**11%**	24	**13%**
4 mentions + P3	6	9%	7	13%	5	8%	18	10%
5 mentions + P3	2	3%	6	11%	3	5%	11	6%
6 mentions + P3	4	6%	3	5%	5	8%	12	6%
7 mentions + P3	0	0%	3	5%	3	5%	6	3%
9 mentions + P3	0	0%	0	0%	2	3%	2	1%
10 mentions + P3	0	0%	2	4%	0	0%	2	1%
17 mentions + P3	0	0%	1	2%	0	0%	1	1%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 45 : Nombre de mentions dans la CDR incluant le référent du SN défini *les enfants*

Les résultats montrent également que, mis à part les cas où P3 est la seule mention du référent, les CDR comportant 1 maillon + celui de P3 (soit 2 maillons au total) sont les plus nombreuses au niveau global : 15% des textes du corpus, bien que leur nombre a tendance à décroître quand le niveau scolaire augmente (23% au CE2, 11% au CM2). En outre, si au niveau global les CDR comportant 3 mentions + P3 sont moins fréquentes (13% du corpus), elles restent majoritaires au CM1 (14%) et au CM2 (11%).

De manière plus minoritaire, les CDR plus longues sont moins nombreuses :

- Entre 4 et 6 mentions + P3, elles représentent entre 18 et 12 textes
- Entre 7 et 10 mentions + P3, elles représentent entre 6 et 1 texte

La longueur moyenne des CDR du SN défini *les enfants* diffère entre le CE2 et les autres niveaux de classe. En effet, si le nombre moyen de maillons est d'environ 2 au CE2 (2,78), il est compris entre environ 3 et 4 dans les autres niveaux : 3,95[56] au CM1 et 3,18 au CM2. La différence constatée peut s'expliquer par les propriétés de P3 qui en font une phrase difficile à traiter en termes de cohérence textuelle. Cette complexité est plus difficile à gérer pour les élèves les plus jeunes, ils prennent donc moins le temps de « dérouler » les CDR des entités qu'elle contient, d'autant plus que l'occurrence *les enfants* comprend déjà en soi plus de difficultés que les « simples » pronoms personnels de P1 et P2.

12.1.2.2. Différentes configurations des CDR incluant le SN défini *les enfants*

Nous l'avons vu, notre corpus regroupe plusieurs configurations de CDR. Il peut s'agir d'anaphore ou de cataphore (1 mention + P3) ou de CDR plus ou moins longues (2 mentions ou plus + P3). Chacune de ces configurations fait l'objet d'une étude plus précise.

Étude des cas d'anaphore : 1 mention + P3

Le cas de figure le plus fréquent dans notre corpus lorsqu'il s'agit de maintenir la référence à l'occurrence *les enfants* est le recours à la configuration suivante : 1 mention + P3 (cf. tableau *supra*). En outre, sur les 15 cas recensés au CE2, 14 sont des cas d'anaphore (la

[56] Pour calculer la longueur moyenne ici nous n'avons pas tenu compte de la CDR de 17 mentions+P3 qui se trouve au CM1 car elle déforme les résultats. En en tenant compte, la longueur moyenne des CDR au CM1 est de 4,27 maillons.

source du référent est localisée avant P3) et 1 seul est un cas de cataphore (la source est placée après P3).

Lorsque la source du référent du SN défini *les enfants* est localisée avant P3, c'est-à-dire dans les cas d'anaphore, on pourrait s'attendre à ce que la première mention soit introduite par un SN indéfini. Or, les cas d'anaphore présentent les mêmes caractéristiques que celles que nous avons rencontrées lors de l'étude de l'introduction du référent (cf. ce chapitre, section 12.1.1.). Pour rappel, il s'agit des particularités de la tâche-problème proposée, qui impliquent que la première mention du référent du SN défini *les enfants* peut être de différentes natures :

- le groupe nominal *les enfants* est constitué par les personnages de P1 et P2, c'est-à-dire les référents des pronoms personnels *elle* et *il*, qui sont réunis à un moment donné dans le texte sous la forme du SN défini *les enfants* ou du pronom personnel *ils*
- le groupe nominal *les enfants* est constitué par des personnages différents de ceux de P1 et P2, intégrés au texte par l'élève sous la forme de SN indéfinis, numéraux ou possessifs
- le groupe nominal *les enfants* réfère à une catégorie d'enfants. Dans ce cas, si le référent est introduit avant P3, il peut se trouver sous la forme d'un SN défini, le défini ayant ici une valeur générique

L'étude de la première mention du référent du SN défini *les enfants* lors de son introduction, c'est-à-dire quel que soit le nombre de maillons de la CDR, a révélé que les SN numéraux étaient la forme la plus fréquente pour introduire la première mention, suivi dans une proportion diminuée de moitié, du SN défini. Dans les cas d'anaphore que nous étudions ici, la première mention du référent, qui est également la seule avant P3, est introduite le plus fréquemment par un SN défini : 44% des cas d'anaphore, comme le montre le tableau 46. Toutefois, cet écart est à relativiser car lorsqu'on regarde le nombre de textes concernés, on constate qu'il y a seulement 2 textes d'écart entre les résultats concernant le SN défini (11 textes) et ceux concernant le SN numéral (9 textes).

Formes de M1 de *les enfants* dans les cas d'anaphore								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN indéfini	1	7%	0	0%	0	0%	1	4%
SN défini	6	43%	2	40%	3	43%	11	**44%**
SN numéral	3	21%	2	40%	4	57%	**9**	36%
SN possessif	1	7%	1	20%	0	0%	2	8%
pronom	3	21%	0	0%	0	0%	2	8%
Total	14	100%	5	100%	7	100%	25	100%

Tableau 46 : Formes de la première mention du référent du SN défini *les enfants* dans les cas d'anaphore (1 mention avant P3 + P3)

L'étude des fonctions de la première mention du référent lors de son introduction (toute longueur de CDR confondue) a montré que la position de sujet syntaxique est plus fréquente que celle de complément de verbe. Dans les cas d'anaphore, c'est le contraire : les textes dans lesquels la première mention est en position de complément de verbe sont plus nombreux (58%) que ceux dans lesquels elle est en position de sujet (31%), comme l'indique le tableau 47.

Fonctions syntaxiques de M1 de *les enfants* dans les cas d'anaphore								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
sujet	6	46%	1	20%	1	14%	8	**31%**
complément de verbe	8	57%	1	20%	6	86%	15	**58%**
construction présentative	0	0%	1	20%	0	0%	1	4%
complément du nom	0	0%	1	20%	0	0%	1	4%
phrase nominale	0	0%	1	20%	0	0%	1	4%
Total	14	100%	5	100%	7	100%	26	100%

Tableau 47 : Fonctions syntaxiques de la première mention du référent du SN défini *les enfants* dans les cas d'anaphore (1 mention avant P3 + P3)

Il semble donc que la longueur de la CDR a une influence sur la nature de ses maillons et leurs fonctions syntaxiques.

Étude des CDR plus longues : 2 mentions (ou plus) + P3

Hormis les cas d'anaphore, dans lesquels il n'y a que deux mentions du référent au total, on trouve des CDR plus ou moins longues dans l'ensemble de notre corpus. Pour rappel, une CDR débute à partir du moment où l'on compte au moins 3 maillons (ou mentions du référent) (Schnedecker, 1997 ; cf. chapitre 2, section 2.2.). Les CDR témoignent d'une gestion de la référence plus globale, d'autant plus lorsqu'elles sont longues. En effet, en mentionnant le référent du SN défini *les enfants* plusieurs fois en dehors de P3, l'élève commence à construire son personnage en lui octroyant une place plus importante.

D'une manière générale dans notre corpus, on constate que plus les CDR comportent de mentions du référent, moins les textes sont nombreux (cf. tableau 45, *supra*) :

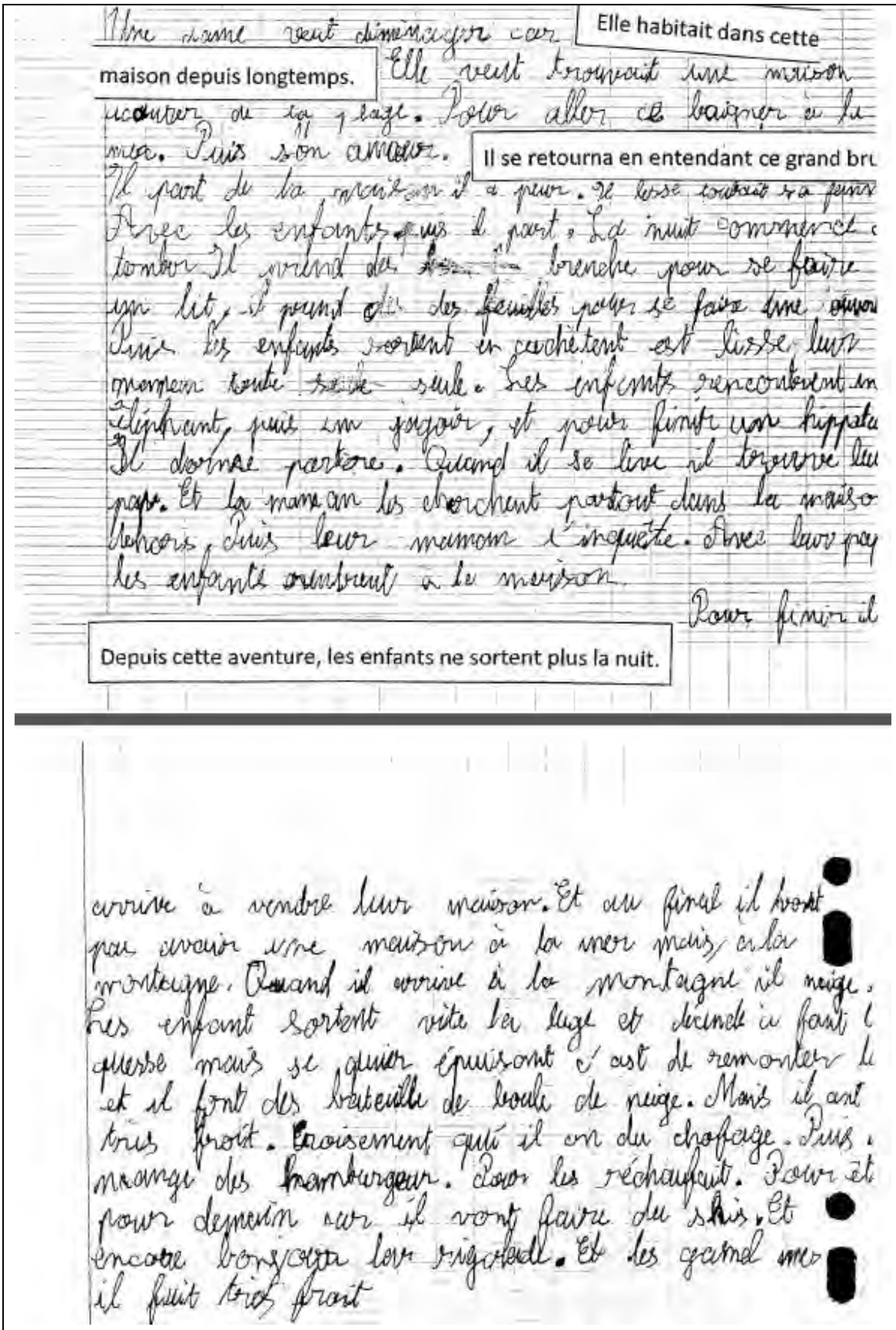
- de 2 à 3 mentions + P3 : entre 15 et 24 textes
- de 4 à 6 mentions + P3 : entre 18 et 12 textes,
- de 7 à 17 mentions + P3 : entre 6 et 1 texte

Nous ne pourrions ici faire une revue détaillée de chacune des mentions du référent du SN défini *les enfants* comme nous l'avons fait jusqu'à présent. Nous nous contenterons donc d'une vue générale des compositions des CDR comportant 2 mentions (ou plus) + P3.

Quel que soit le nombre de maillons de la CDR incluant le SN défini *les enfants,* et quel que soit le niveau de classe, le pronom est la forme la plus fréquente pour maintenir la référence. Ce résultat est en accord avec les travaux qui se sont intéressés au maintien de la référence chez les enfants ou les adultes, notamment les travaux de Jisa (2000 ; 2004 ; cf. chapitre 3, section 3.1.2.2). Ses recherches avec des enfants de 5, 7 et 10 ans ont confirmé la prédominance de l'emploi des pronoms dans le maintien de la référence, quel que soit l'âge des enfants. Bien que portant sur l'oral, la situation de connaissance non partagée qui a été testée se rapproche de nos propres travaux sur l'écrit.

L'étude des fonctions syntaxiques montre que quelle que soit la longueur de la CDR incluant le SN défini *les enfants*, le référent est exclusivement maintenu en position de sujet syntaxique.

Nous retrouvons ces éléments dans le texte d'Elouan (CM1 T1), que nous reproduisons dans son intégralité et qui comporte la plus longue CDR incluant le SN défini *les enfants* dans notre corpus, puisqu'elle comporte 17 mentions du référent + P3 (soit 18 mentions au total).



Une dame veut déménager car *elle habitait dans cette maison depuis longtemps*. Elle veut trouver une maison à côté de la plage. Pour aller se baigner à la mer. Puis son amour. *Il se*

retourna en entendant ce grand bruit. Il part de la maison il a peur. Il laisse tomber sa femme, avec les enfants. Puis il part. La nuit commence à tomber. Il prend des branches pour se faire un lit, il prend des feuilles pour se faire une couverture. Puis les enfants sortent de leur cachette et laissent leur maman toute seule. Les enfants rencontrent un éléphant, puis un jaguar, et pour finir un hippopotame. Ils dorment par terre. Quand ils se lèvent, ils trouvent leur papa. Et la maman les cherche partout dans la maison, dehors. Puis leur maman s'inquiète. Avec leur papa, les enfants rentrent à la maison. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Pour finir, il arrive à vendre leur maison. Et au final, ils ne vont pas avoir une maison à la mer mais à la montagne. Quand ils arrivent à la montagne, il neige. Les enfants sortent vite la luge et descendent à fond la caisse mais ce qui est épuisant, c'est de remonter (la pente) et ils font des batailles de boule de neige. Mais ils ont très froid. Heureusement qu'ils ont du chauffage. Puis ils mangent des hamburgers. Pour les réchauffer. Peut-être pour demain, sûr, ils vont faire du ski. Et encore, bonjour leurs rigolades. Et les gamelles ! Mais il fait très froid.

Figure 115 : Texte d'Elouan, CM1 T1

Dans ce texte, le SN défini *les enfants* réfère à des personnages qui ne sont pas ceux de P1 et P2 mais des enfants insérés dans l'histoire, qui sont présentés en groupe tout au long du texte. Le référent du SN défini *les enfants* est mentionné 9 fois avant P3. La première mention est sous la même forme que dans P3, il s'agit du SN défini *les enfants* qui désigne les enfants du couple que forment les référents des pronoms *elle* et *il* de P1 et P2. D'abord placé en position de complément du nom *femme*, les personnages des enfants sont réinstanciés quelques phrases plus loin à l'aide du même SN défini, placé cette fois en position de sujet syntaxique. La fonction sujet place le référent dans le focus d'attention du lecteur et en fait le sujet thématique pour plusieurs phrases, en témoignent les pronoms (personnels et zéro) employés pour maintenir la référence. C'est la stratégie du sujet thématique (Karmiloff-Smith, 1980 ; 1981 ; cf. chapitre 3, section 3.1.2.2.).

Après P3, le texte continu. Le référent du SN défini *les enfants* est mentionné 8 fois après P3. Sorti du focus d'attention du lecteur, il est de nouveau réinstancié à l'aide du SN défini *les enfants* en position de sujet syntaxique dans la phrase *Les enfants sortent vite la luge…* et la référence est maintenue à l'aide de pronoms, principalement en position de sujet syntaxique, jusqu'à la fin du texte, conformément à ce qui est attendu.

12.2. Le SN défini *la nuit*

Dans la troisième phrase imposée de la consigne (P3), *Depuis cette aventure, les enfants ne sortent plus la nuit*, le SN défini *la nuit* est une indication temporelle à prendre en compte dans le texte et nécessite l'introduction préalable d'un référent. En 12.2.1., nous nous attacherons donc à décrire les diverses formes de la première mention du référent, ses éventuelles expansions ainsi que les fonctions syntaxiques qu'il peut occuper. En 12.2.2., nous étudierons comment les élèves maintiennent le référent tout au long du discours.

12.2.1. Introduction du référent du SN défini *la nuit*

P3 étant la dernière phrase à insérer dans le récit, sa valeur de phrase *fermoir* (Marandin, 1986) implique que l'indication temporelle *la nuit* soit prise en compte dans l'élaboration de la totalité du texte. Outre la nécessité, sur le plan sémantique, de ne pas contenir de contradiction de monde (méta-règle de non contradiction, Charolles, 1978 ; cf. chapitre 1, section 1.1.2.) qui situerait l'action du récit pendant la journée alors que P3 la situe la nuit, sur le plan linguistique le référent du SN défini *la nuit* nécessite une introduction préalable à P3, au même titre que les autres occurrences de la consigne. Cependant, le SN défini *la nuit* étant une indication temporelle, on ne s'attend pas à ce que la référence soit maintenue très longtemps dans le texte.

Aussi, on observe deux localisations possibles de l'expression référentielle qui introduit le référent (ou *source*), comme le montre le tableau 48 :

- La source (ou référent) est localisée avant la phrase imposée
- La source est absente : la seule mention du référent est celle de la phrase imposée

Localisation de la source de *la nuit*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
avant P3	38	59%	27	48%	34	52%	99	**54%**
source absente	26	41%	29	52%	31	48%	86	**46%**
Total	64	100%	56	100%	65	100%	185	100%

Tableau 48 : Localisation de la source du SN défini *la nuit*

Les résultats de notre étude montrent que la majorité des textes localisent la source du référent avant P3 : 54% des élèves de notre corpus. Mais les textes dans lesquels il n'y a aucune mention du référent mise à part celle contenue dans P3 sont également très nombreux : ils représentent 46% du corpus. Globalement, on peut donc dire qu'une moitié des textes traite l'indication temporelle et l'autre moitié ne le fait pas. Ce résultat peut sans doute s'expliquer premièrement par la position de l'occurrence *la nuit* dans la consigne. En effet, dans l'ordre des phrases imposées, c'est la dernière occurrence à traiter. Deuxièmement, le caractère plus secondaire de cette indication temporelle peut la faire passer au second plan devant les autres occurrences de la consigne à traiter, plus saillantes.

L'analyse plus détaillée des résultats par niveau scolaire met en lumière un fait surprenant. En effet, on pourrait penser que l'indication temporelle *la nuit* présente un caractère « secondaire » comparé à toutes les occurrences de la consigne à traiter. Ce sont donc les élèves les plus jeunes qui devraient avoir le plus de mal à traiter cette occurrence dans leur texte, d'autant plus qu'elle nécessite une planification à rebours (Hayes & Flower, 1980 ; cf. chapitre 3, section 3.1.4.). Or, les résultats de notre étude montrent que les élèves les plus jeunes sont plus nombreux à introduire le référent du SN défini *la nuit* : 59% au CE2 contre 48% au CM1 et 52% au CM2. De la même manière, ce sont les élèves les plus jeunes qui ont le taux le moins élevé d'absence de source du référent : 41% au CE2 contre 52% au CM1 et 48% au CM2.

Après avoir dressé une vue d'ensemble des différentes manières de gérer la référence, nous allons examiner plus en détail les formes de la première mention du référent du SN défini *la nuit.*

Les différentes formes d'introduction du référent du SN défini **la nuit**

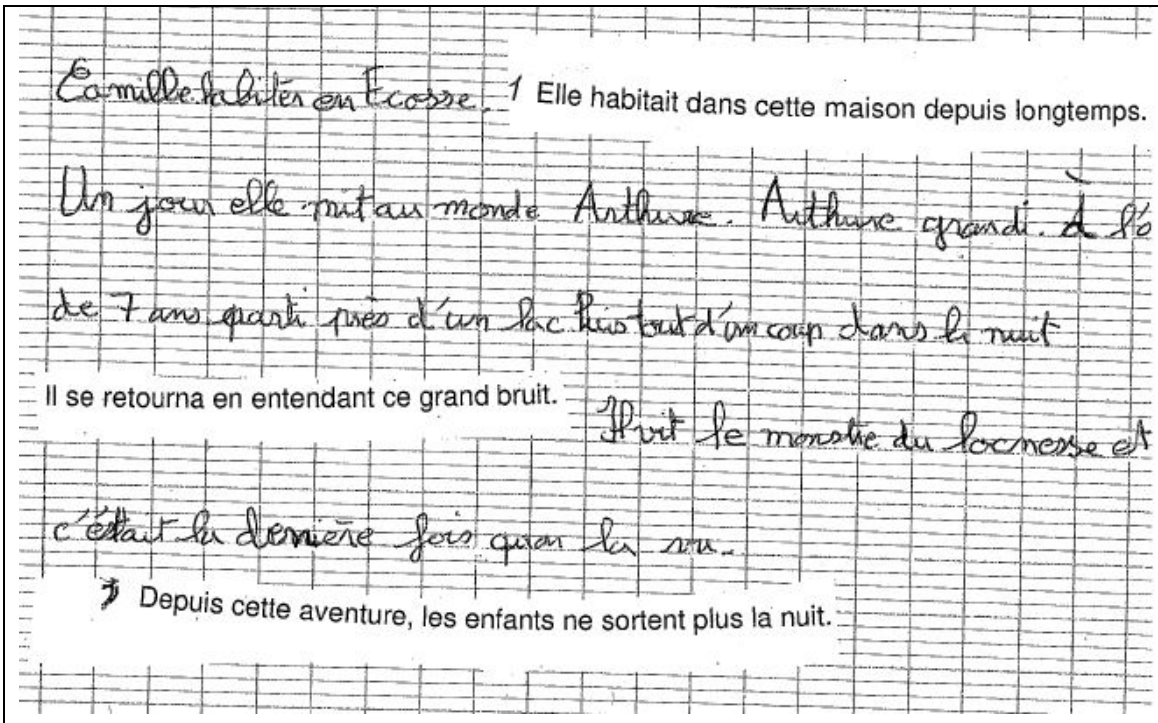
Le SN défini *la nuit* est une indication temporelle dont le sémantisme du nom tête, *nuit*, lui permet de se réaliser sous diverses formes. En effet, en tant que mot « plein » (Mortureux, 1998), le nom *nuit* possède une certaine valeur dénominative qui lui permet d'évoquer la réalité qui lui est associée. De ce fait, outre la reprise du nom *nuit* accompagné de différents déterminants dans les SN, le référent de l'occurrence *nuit* peut également se trouver sous la forme de phrases complètes nécessitant ou non une inférence.

Le tableau 49 reproduit les différents types de formes de la première mention du référent du SN défini *la nuit* et leur fréquence d'emploi dans notre corpus.

Formes de M1 pour *la nuit*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
SN défini	22	34%	13	23%	20	31%	55	**30%**
SN indéfini	7	11%	9	16%	6	9%	22	**12%**
SN démonstratif	3	5%	2	4%	1	2%	6	3%
phrase complète	5	8%	3	5%	7	11%	15	8%
P3 = seule mention	27	42%	29	52%	31	48%	87	47%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 49 : Formes de la première mention du référent du SN défini *la nuit*

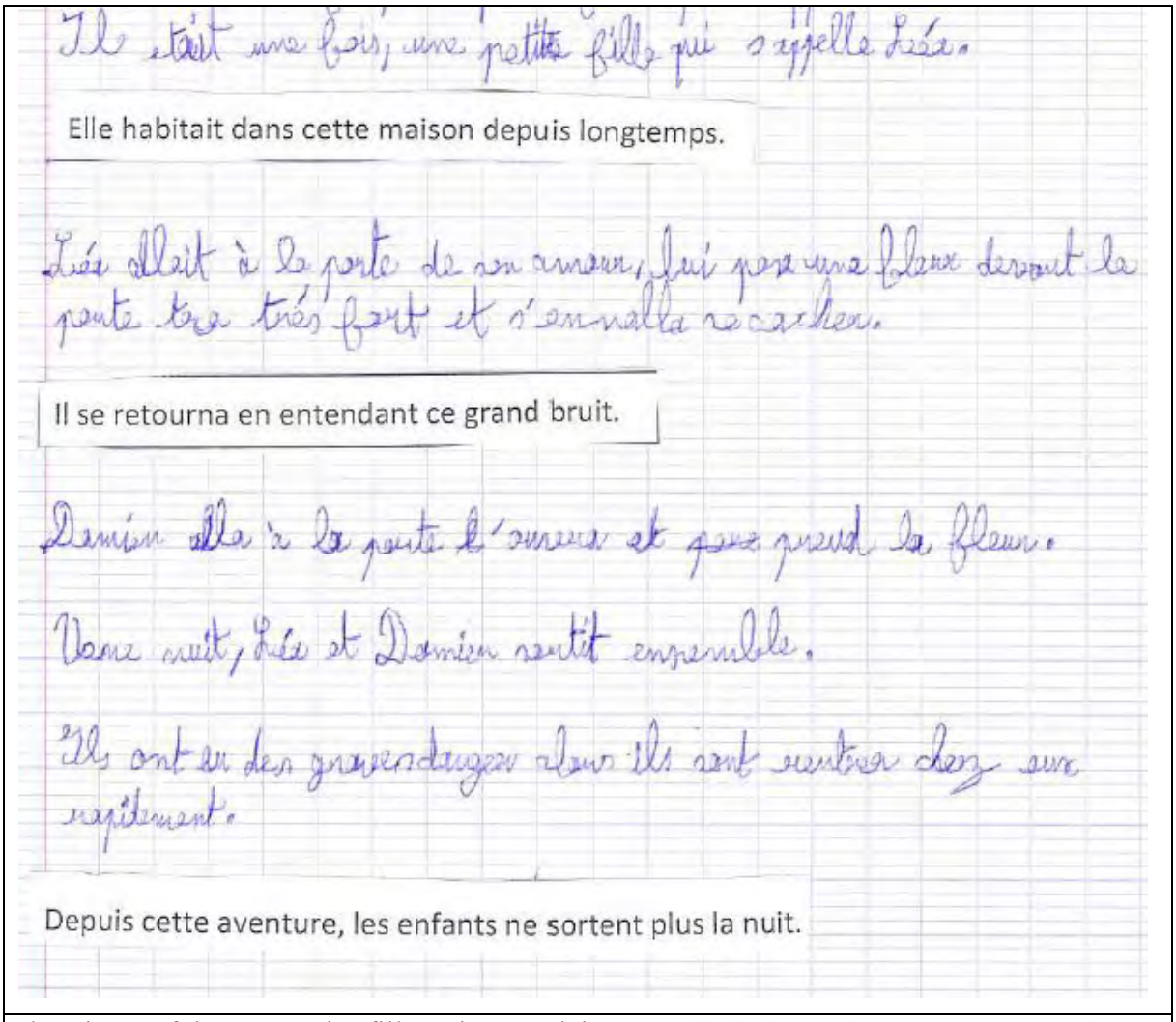
Nous l'avons déjà vu, l'introduction d'un référent se fait préférentiellement sous la forme d'un SN indéfini (cf. ce chapitre, section 12.1.1.1.). Cependant, par sa position dans la phrase *fermoir* P3, le SN défini *la nuit* revêt un caractère générique (cf. ce chapitre, introduction). Il est donc naturel d'employer un SN défini (dans son sens générique) comme forme de première mention du référent. C'est d'ailleurs ce que font principalement les élèves de notre corpus : 30% des textes introduisent le référent sous la forme d'un SN défini, comme dans (116) :



Camille habitait en Ecosse. *Elle habitait dans cette maison depuis longtemps.* Un jour, elle mit au monde Arthur. Arthur grandit. À l'âge de 7 ans, partit près d'un lac. Puis tout à coup dans la nuit, *il se retourna en entendant ce grand bruit.* Il vit le monstre du Loch Ness et c'était la dernière fois qu'on l'a vu. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 116 : Texte d'Enzo, CM2 S

La deuxième forme de première mention la plus fréquente est le SN indéfini : 12% dans l'ensemble du corpus. Mais l'emploi du SN indéfini ici sort de son cadre habituel. En effet, les SN indéfinis sont habituellement employés en première mention car ils sont autonomes référentiellement parlant (Charolles, 2002 ; cf. chapitre 1, section 1.4.2.1.) : ils ne demandent pas à ce que l'interlocuteur sache déjà de quoi il va être question dans le propos. En cela, ils permettent généralement de fixer le référent et non de le désigner (comme c'est le cas des SN définis et démonstratifs). Dans le cas de l'occurrence *la nuit,* employée dans un sens générique, le SN indéfini *une nuit* permet justement de désigner une nuit spécifique, de la faire « sortir du lot » de toutes les autres nuits. La valeur du SN indéfini employé dans notre cas se rapproche donc de la valeur ordinaire des SN définis et démonstratifs. On le voit dans (119), où le SN indéfini *une nuit* réfère à une nuit en particulier : celle où Léa et Damien sont sortis ensemble.



Il était une fois, une petite fille qui s'appelle Léa.

Elle habitait dans cette maison depuis longtemps.

Léa allait à la porte de son amour, lui posa une fleur devant la porte toqua très fort et s'en alla se cacher.

Il se retourna en entendant ce grand bruit.

Damien alla à la porte l'ouvra et posa prend la fleur.

Une nuit, Léa et Damien sortit ensemble.

Ils ont eu des graves dangers alors ils sont rentrer chez eux rapidement.

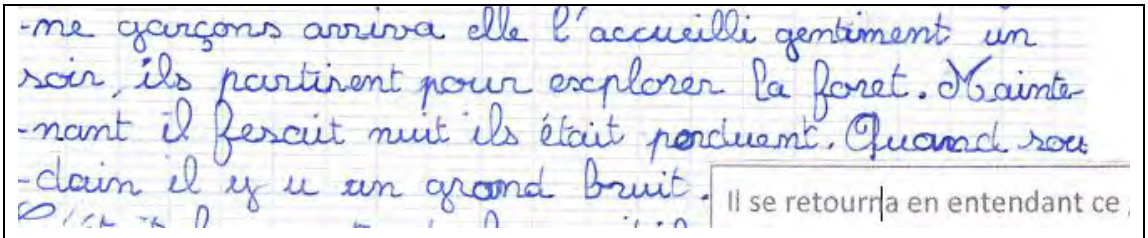
Depuis cette aventure, les enfants ne sortent plus la nuit.

Il était une fois, une petite fille qui s'appelait Léa. *Elle habitait dans cette maison depuis longtemps.* Léa allait à la porte de son amour, lui posa une fleur devant la porte, toqua très fort et s'en alla se cacher. *Il se retourna en entendant ce grand bruit.* Damien alla à la porte, l'ouvrit et prit la fleur. Une nuit, Léa et Damien sortirent ensemble. Ils ont eu des graves dangers alors ils sont rentrés chez eux rapidement. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 117 : Texte d'Ange, CE2 P

L'étude des formes de la première mention du référent du SN défini *la nuit* montre également que 8% des textes du corpus contiennent une phrase entière pour référer au SN défini *la nuit.* Nous pouvons distinguer trois types de situations :

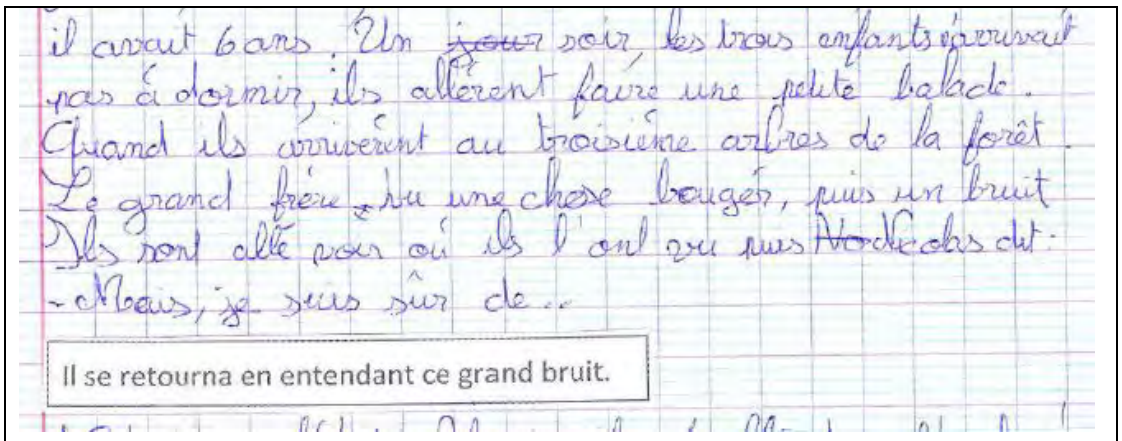
- des phrases qui contiennent le mot *nuit* dans une expression prédicative comme dans (118) avec *il faisait nuit* :


-me garçons arriva elle l'accueilli gentiment un soir, ils partirent pour explorer la foret. Mainte-nant il fesait nuit ils était perduent. Quand sou-dain il y u un grand bruit. Il se retourna en entendant ce

… un soir, ils partirent pour explorer la forêt. Maintenant il faisait nuit, ils étaient perdus. Quand soudain, il y eut un grand bruit. *Il se retourna en entendant ce grand bruit.*

Figure 118 : Texte de Juliette, CE2 T

- des phrases qui contiennent un ou plusieurs éléments lexicaux dont le sémantisme peut les inclure dans le champ notionnel de *nuit*, comme par exemple le SN *un soir* ou le verbe *dormir* dans (119) :


il avait 6 ans. Un ~~jour~~ soir, les trois enfants n'arrivait pas à dormir, ils allèrent faire une petite balade. Quand ils arrivèrent au troisième arbres de la forêt. Le grand frère a vu une chose bouger, puis un bruit. Ils sont allé voir où ils l'ont vu puis Nicolas dit : - Mais, je suis sûr de..
Il se retourna en entendant ce grand bruit.

… Un soir, les trois enfants n'arrivaient pas à dormir, ils allèrent faire une petite balade. Quand ils arrivèrent au troisième arbre de la forêt. Le grand-frère vit une chose bouger, puis un bruit. Ils sont allés voir où ils l'ont vu puis Nicolas dit :

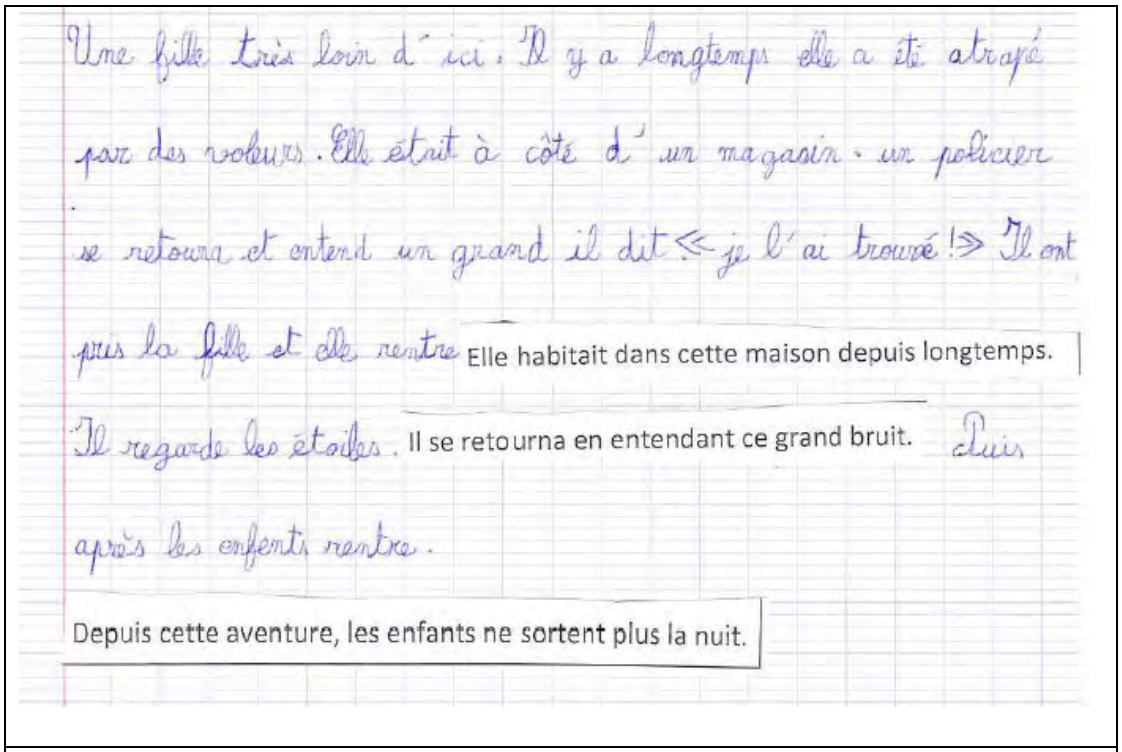
- Mais, je suis sûr de…

Il se retourna en entendant ce grand bruit.

Figure 119 : Texte de Solène, CM1 T2

- des phrases qui ne contiennent ni le mot *nuit* ni aucun autre élément lexical dont le sémantisme à lui seul peut l'inclure dans le champ notionnel de *nuit*. Le lecteur doit

réaliser un calcul inférentiel pour comprendre que l'action se déroule la nuit. C'est le cas par exemple dans (120) :


Une fille très loin d'ici. Il y a longtemps elle a été atrapé par des voleurs. Elle était à côté d'un magasin. un policier se retourna et entend un grand il dit « je l'ai trouvé ! » Il ont pris la fille et elle rentre Elle habitait dans cette maison depuis longtemps. Il regarde les étoiles. Il se retourna en entendant ce grand bruit. Puis après les enfants rentre.

Depuis cette aventure, les enfants ne sortent plus la nuit.

Une fille très loin d'ici. Il y a longtemps, elle a été attrapée par des voleurs. Elle était à côté d'un magasin. Un policier se retourna et entend un grand (bruit ?), il dit : « je l'ai trouvée ! ». Ils ont pris la fille et elle rentre. *Elle habitait dans cette maison depuis longtemps.* Il regarde les étoiles. *Il se retourna en entendant ce grand bruit.*

Puis après les enfants rentrent. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 120 : Texte d'Enzo, CE2 P

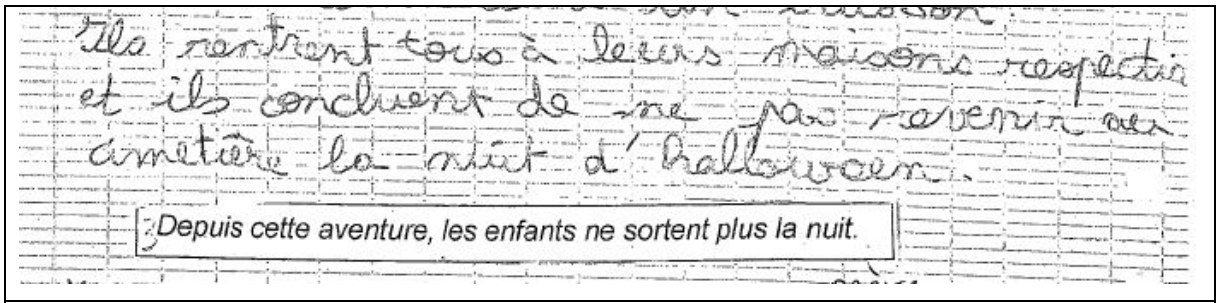
Dans ce texte, le sémantisme du mot *étoile* suffit à lui seul à faire comprendre que l'action se déroule la nuit étant donné que les étoiles ne sont visibles que la nuit. Nous précisons que ce texte fait l'objet d'une analyse plus qualitative dans la section 12.3. étant donné le manque de cohérence textuelle globale.

Les expansions de la première mention du référent du SN défini la nuit

Lors de l'introduction du référent du SN défini *la nuit* présent dans P3, des SN peuvent être accompagnés d'un modifieur qui est autant d'ajout d'information à propos de l'entité ainsi dénotée (Charolles, 2002).

L'étude des formes de la première mention du référent du SN défini *la nuit* montre que si, dans leur grande majorité, les textes du corpus ne présentent pas d'expansion (95%), on en trouve tout de même 9 dans lesquels la forme de la première mention est expansée, principalement par un syntagme prépositionnel (6 textes) qui donne une indication précise telle que *à minuit* ou *la nuit du 17 mars,* mais également par un syntagme adjectival (3 textes) qui situe la nuit en question par rapport à d'autres : *suivante, précédente, deuxième*.

On remarque également que dans les classes de CM2, le syntagme prépositionnel est le seul type d'expansion employé, comme dans l'exemple (121) avec le SN expansé *la nuit d'Halloween* :


Ils rentrent tous à leurs maisons respectives et ils concluent de ne pas revenir au cimetière la nuit d'Halloween.

Depuis cette aventure, les enfants ne sortent plus la nuit.

… Ils rentrent tous à leur maison respective et ils concluent de ne pas revenir au cimetière la nuit d'Halloween. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 121 : Texte de Salomé, CM2 B

***Les fonctions syntaxiques de la première mention du référent du SN défini* la nuit**

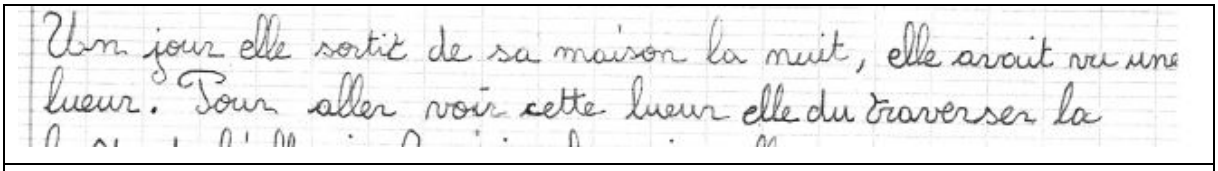
Nous l'avons déjà montré, lorsque la première mention d'un référent apparaît, c'est une information totalement nouvelle (cf. *statut de l'information*, chapitre 1, section 1.2.) qui représente le thème du propos (cf. *progression thématique*, chapitre 1, section1.2.3.). Dans ce cas, il est naturel que cette nouvelle information prenne la place de sujet syntaxique et se retrouve en position initiale dans l'énoncé.

Cependant, le caractère « secondaire » de l'occurrence *nuit*, dû à sa nature d'indication temporelle, place le SN défini *la nuit* de P3 en position de complément de phrase. Les élèves auront tendance à garder la même position syntaxique que celle de P3 et le référent du SN défini devrait alors se retrouver en position de complément de phrase, plus rarement en position de sujet syntaxique.

Cette hypothèse est confirmée par nos résultats : 35% des élèves du corpus emploient la première mention du référent du SN défini *la nuit* en position de complément de phrase, comme l'illustrent le tableau 50 et l'exemple (122) ci-après.

Fonctions syntaxiques de M1 pour *la nuit*								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
complément de phrase	26	41%	16	29%	23	35%	65	**35%**
complément de verbe	1	2%	3	5%	1	2%	**5**	3%
sujet	0	0%	5	9%	3	5%	8	4%
phrases entières	5	8%	3	5%	7	11%	15	8%
P3 = seule mention	32	50%	29	52%	31	48%	92	50%
Total	64	100%	56	100%	65	100%	185	100%

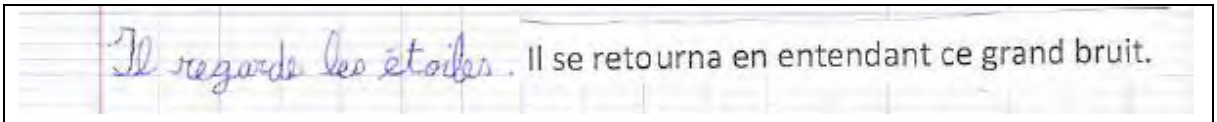
Tableau 50 : Types de fonctions syntaxiques prises par la première mention du référent du SN défini *la nuit*



Un jour elle sortit de sa maison la nuit, elle avait vu une lueur. Pour aller voir cette lueur elle dû traverser la

Figure 122 : Texte de Quentin, CM2 M

De manière plus marginale, on constate quelques cas où le référent est placé en position de complément de verbe (5 textes). Il s'agit de cas où le référent du SN défini *la nuit* ne contient pas la même tête lexicale que dans P3. Dans (122) que nous avons déjà analysé *supra* mais que nous reprenons, le référent du SN défini *la nuit* est le SN défini *les étoiles*, identifiable par inférence puisque les étoiles ne sont visibles que la nuit :



Il regarde les étoiles. *Il se retourna en entendant ce grand bruit.*

Figure 123 : Texte d'Enzo, CE2 P

12.2.2. Maintien du référent du SN défini *la nuit*

Le statut d'indication temporelle de l'occurrence *nuit* en fait une occurrence qu'il n'est pas nécessaire de répéter plus de 2 fois dans le discours : la mention contenue dans P3 et son introduction peuvent suffire, à moins bien entendu de raconter une histoire dans laquelle *la nuit* est le sujet thématique, ce qui ne semble pas être le cas dans notre étude.

Les résultats de notre recherche confirment cette hypothèse. En effet, si 47% des textes de notre corpus ne contiennent pas l'occurrence *nuit* en dehors de P3, dans 51% des textes il y a une anaphore pour mentionner le référent (1 mention + P3, soit deux mentions au total).

Les textes dans lesquels il y a une véritable CDR, c'est-à-dire à partir de trois maillons (Schnedecker, 1997 ; cf. chapitre 2, section 2.2.), sont très peu nombreux : 1 texte contient une CDR de 3 mentions (2 mentions + P3) et 1 texte contient une CDR de 4 mentions (3 mentions + P3). Nous reproduisons ci-dessous dans son intégralité le texte qui comporte la plus longue CDR incluant le SN défini *la nuit* dans notre corpus, avec deux mentions du référent + celle de P3 : le SN indéfini *une nuit* contenu dans la proposition subordonnée temporelle *Quand une nuit*, et la phrase *ils sortirent c'était la nuit* :

Camille était une jeune fille aux yeux bleus et avait des cheveux blonds. Cette fille vivait dans une maison. Elle habitait dans cette maison depuis longtemps.

Cette enfant avait ses parents, un frère et un chat. Son frère nommé Damien vivait lui aussi depuis longtemps dans cette affreuse maison. On aurait dit une maison hantée. Quand une nuit, il entenda un bruit énervant qui faisait peur. Un bon moment. Il se retourna en entendant ce grand bruit. Il sortit de sa chambre.

Tout le monde dormait. Pas un bruit. Alors, il alla réveillait sa soeur. Ils sortirent c'était la nuit. Ils entendirent du bruit dans les buissons, c'était la pleine lune, il était exactement 00h00. Les deux enfants virent des ragondins en train de manger leur petit chat. Ils se dépêchèrent de revenir dans leur maison avec leurs parents. Depuis cette aventure, les enfants ne sortent plus la nuit.

Camille était une jeune fille aux yeux bleus et avait des cheveux blonds. Cette fille vivait dans une maison. *Elle habitait dans cette maison depuis longtemps.* Cette enfant avait ses parents, un frère et un chat. Son frère nommé Damien vivait lui aussi depuis longtemps dans cette affreuse maison. On aurait dit une maison hantée. Quand une nuit, il entendit un bruit énervant qui faisait peur. Un bon moment. *Il se retourna en entendant ce grand bruit.* Il sortit de sa chambre. Tout le monde dormait. Pas un bruit. Alors, il alla réveiller sa sœur. Ils sortirent c'était la nuit. Ils entendirent du bruit dans les buissons, c'était la pleine lune, il était exactement 00h00. Les deux enfants virent des ragondins en train de manger leur petit chat. Ils se dépêchèrent de revenir dans leur maison avec leurs parents. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 124 : Texte de Noëmie, CM2 B

12.3. Enchâssement de P3 et résolution de l'anaphore résomptive *depuis cette aventure*

Dans la tâche-problème proposée, nous l'avons vu, P3 n'est pas une phrase comme les autres. En plus de la prise en compte des référents des SN définis *les enfants* et *la nuit*, à la fois sur les plan sémantique et syntaxique, les élèves sont également amenés à résoudre l'anaphore résomptive *depuis cette aventure*, qui nécessite une planification à rebours de leur texte.

Comme nous l'avons vu au chapitre 2 (section 2.1.2.1.), l'anaphore résomptive est un cas particulier d'anaphore. En effet, contrairement à la plupart des anaphores, l'anaphore résomptive se manifeste par une expression démonstrative à travers laquelle le référent visé n'est pas un élément linguistique mais une représentation mentale. Son objectif n'est pas de référer à une relation de coréférence, une entité « objective » du discours (un personnage par exemple) mais à un contenu posé ou présupposé et d'instituer un cadre interprétatif (Kara & Wiederspiel, 2011). Pour illustrer ces propos, nous reprenons un exemple de Himmelman (1996) cité par Guillot (2006) :

(119) … puis il part… et c'est la fin de *cette histoire*, mais après… il revient vers le paysan…

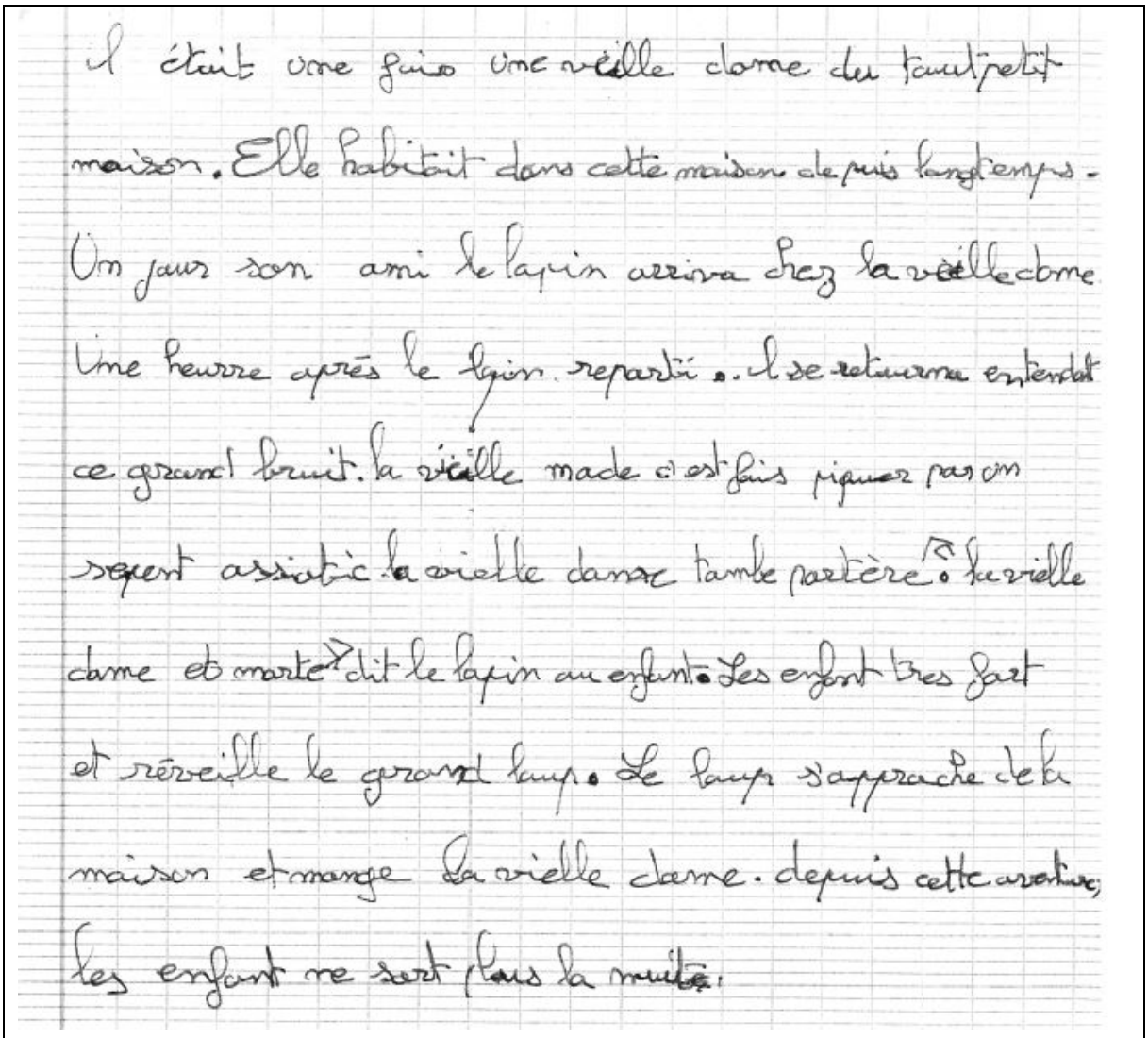
Dans cet exemple, le SN démonstratif *cette histoire* est anaphorique car les informations contenues dans le contexte précédent sont nécessaires à son interprétation. Toutefois, ce type d'anaphore réfère au contenu discursif d'un ensemble de propositions venant d'être énoncées (de quelques propositions à un discours entier, Maillard, 1974), sans que ce contenu n'ait été désigné auparavant par une expression référentielle particulière (Guillot, 2006) : ce sont les propositions énoncées antérieurement qui font de ce discours une histoire.

De la même manière avec l'anaphore résomptive de notre consigne, *depuis cette aventure*, ce sont les propositions énoncées avant P3 dans le texte des élèves qui contribuent à en faire un récit que l'on peut qualifier d'aventure, et qui permettent « d'accrocher » P3 à l'ensemble du texte de manière à ce que la cohérence textuelle soit respectée.

Aussi, pour que l'enchâssement de P3 dans le texte soit réussi, il faut satisfaire trois critères :

- le référent du SN défini *les enfants* est localisable dans le texte : par inférence dans les cas où P3 est la seule mention, ou bien par une expression linguistique appropriée que l'élève peut choisir de maintenir ou non dans son texte (CDR plus ou moins longue)
- l'indication temporelle *la nuit* est présente dans le récit : sur le plan sémantique par une action située la nuit et/ou sur le plan linguistique par l'emploi d'une expression appropriée qui réfère au SN défini *la nuit*
- l'anaphore résomptive *depuis cette aventure* est résolue : les évènements racontés en amont dans le texte peuvent être qualifiés d'aventure et conduire à l'abandon des sorties nocturnes des enfants

Pour des apprentis scripteurs, ces critères ne sont pas nécessairement faciles à satisfaire tous en même temps. Pour cette raison, beaucoup de textes n'en satisfont qu'un ou deux sur les trois, comme (124) :



il était une fois une vieille dame du tout petit
maison. Elle habitait dans cette maison depuis longtemps.
Un jour son ami le lapin arriva chez la vieille dame.
Une heure après le lapin reparti. Il se retourna en entendant
ce grand bruit. la vieille made c'est fais piquer par un
serpent assiatic la vieille danse tombe partère. la vieille
dame est morte dit le lapin au enfant. Les enfant très fort
et réveille le grand loup. Le loup s'approche de la
maison et mange la vieille dame. depuis cette aventure,
les enfant ne sort plus la nuit.

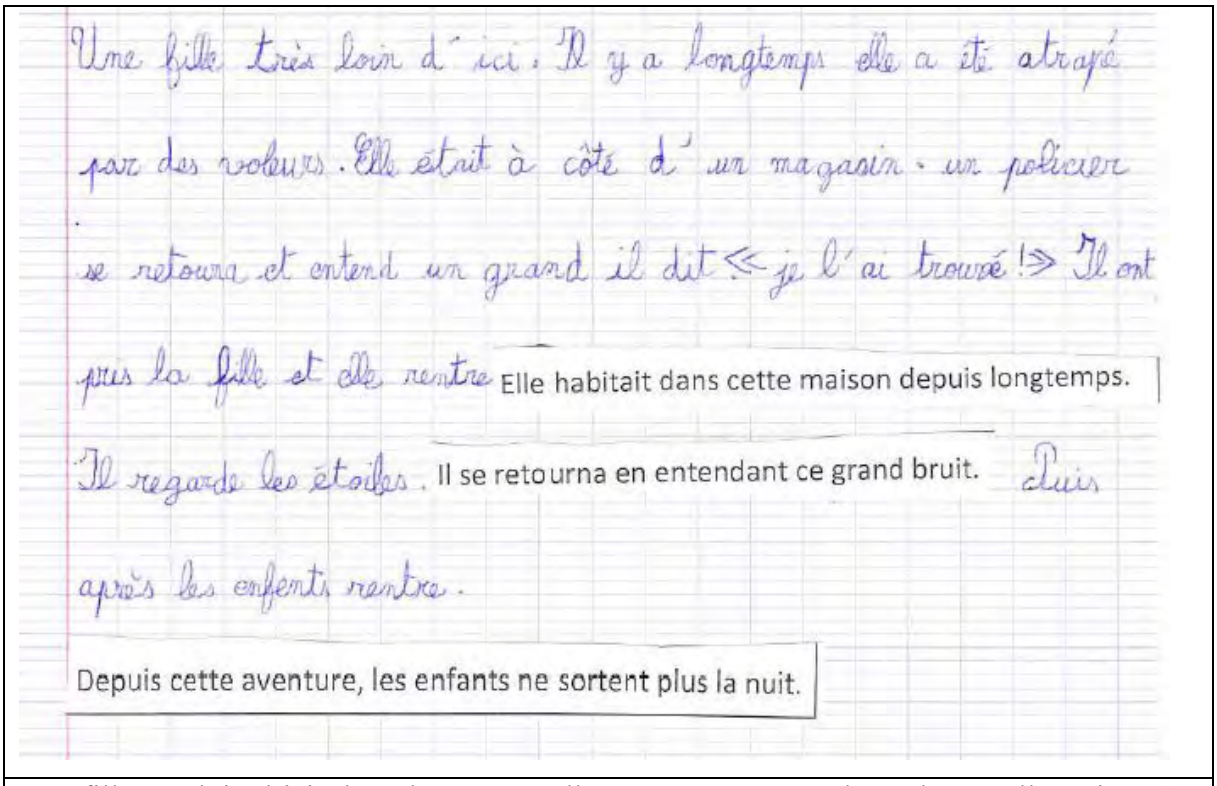
Il était une fois une vieille dame (dans une) toute petite maison. *Elle habitait dans cette maison depuis longtemps.* Un jour, son ami le lapin arriva chez la vieille dame. Une heure après, le lapin reparti. *Il se retourna en entendant ce grand bruit.* La vieille dame s'est fait piquer par un serpent asiatique, la vieille dame tombe par terre. « La vieille dame est morte », dit le lapin aux enfants. Les enfants (crient ? pleurent ?) très fort et réveillent le grand loup. Le loup s'approche de la maison et mange la vieille dame. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 125 : Texte d'Ugo, CM2 M

Dans cet exemple, l'indication temporelle *la nuit* n'est pas présente : à aucun moment dans le texte on ne peut comprendre que la scène se déroule pendant la nuit. En outre, si linguistiquement parlant, le référent du SN défini *les enfants* est introduit avant P3 par le SN défini *aux enfants*, du point-de-vue sémantique, on ne comprend pas de quels enfants il s'agit étant donné qu'il n'y a pas de candidat à la référence en amont dans le texte. Enfin, l'anaphore

résomptive n'est pas vraiment résolue. D'un côté, le récit peut être qualifié d'aventure par le fait qu'il y a la présence d'un danger et d'une peur (le loup qui mange les humains). Mais de l'autre, on comprend mal comment le fait que le loup mange le cadavre de la vieille dame peut avoir un impact aussi fort sur les enfants, les conduisant à ne plus *sortir*. Il aurait été peut-être plus acceptable pour un lecteur que le fait que le loup mange la vieille dame conduisent les enfants à ne plus *crier* ou *pleurer*, mais pas à ne plus *sortir*. Remarquons tout de même que ce dernier point relève du degré de coopération plus ou moins grand de chaque lecteur et comporte donc une part de subjectivité.

Parfois même, tous les critères semblent satisfaits, mais on ne peut pour autant conclure à un enchâssement correct de P3 dans le texte, comme dans (125) :

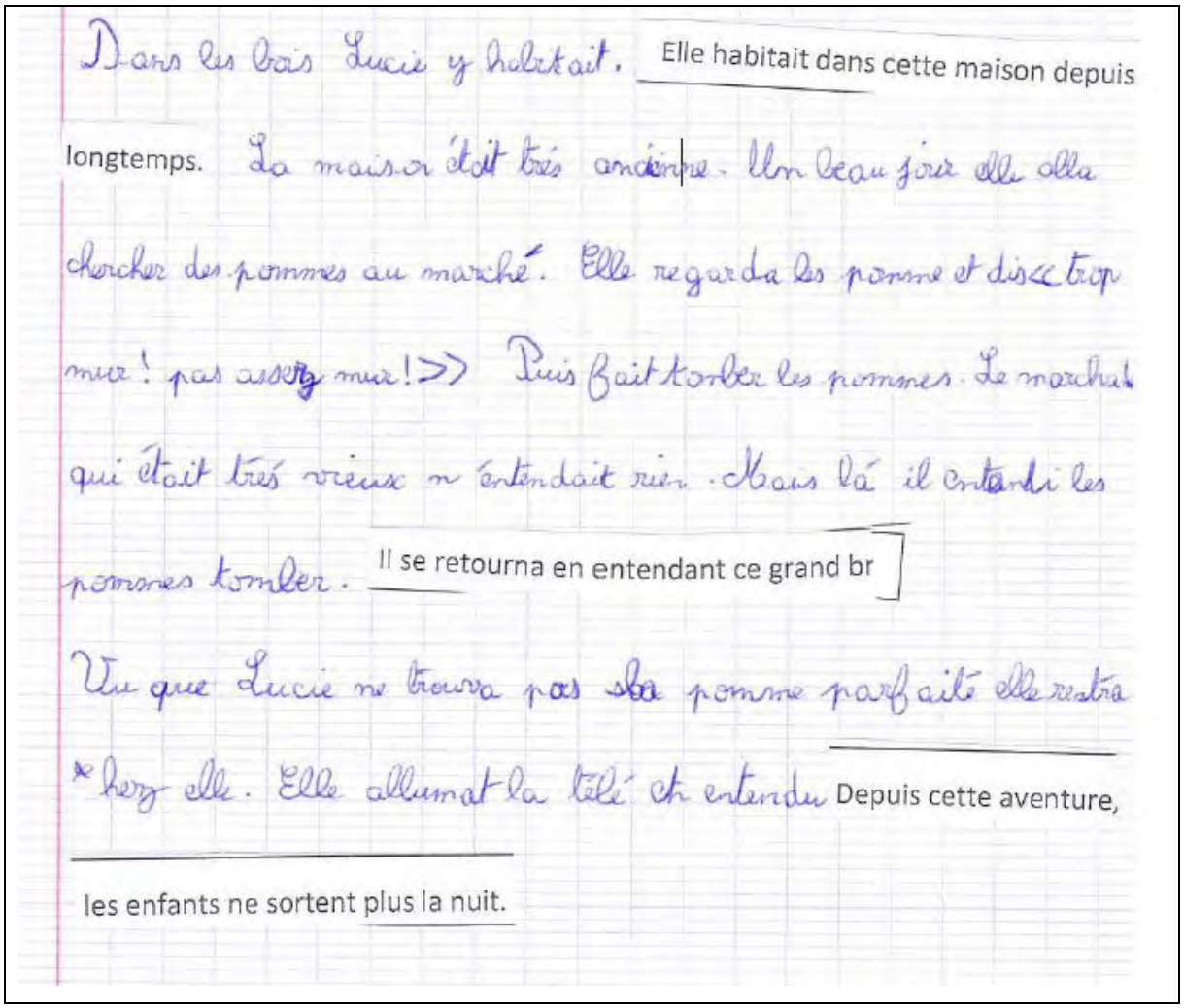

Une fille très loin d'ici. Il y a longtemps elle a été atrapé
par des voleurs. Elle était à côté d'un magasin. un policier
se retoura et entend un grand il dit « je l'ai trouvé ! » Il ont
pris la fille et elle rentre Elle habitait dans cette maison depuis longtemps.
Il regarde les étoiles. Il se retourna en entendant ce grand bruit. Puis
après les enfants rentre.
Depuis cette aventure, les enfants ne sortent plus la nuit.

Une fille très loin d'ici. Il y a longtemps elle a été attrapée par des voleurs. Elle était à côté d'un magasin. Un policier se retourna et entend un grand (bruit ?), il dit : « je l'ai trouvé ! ». Ils ont pris la fille et elle rentre. *Elle habitait dans cette maison depuis longtemps.* Il regarde les étoiles. *Il se retourna en entendant ce grand bruit.* Puis les enfants rentrent. *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Figure 126 : Texte d'Enzo, CE2 P

Dans cet exemple, le SN défini *les enfants* est introduit juste avant P3 par la même expression linguistique, l'indication temporelle *la nuit* est prise en compte à travers l'inférence entraînée par le SN défini *les étoiles* et le début du récit met bien en scène une aventure (enlèvement d'une fille et retrouvailles). Pourtant, le lecteur a du mal à « raccrocher » P3 au texte. Il semble que la méta-règle de relation (Charolles, 1978 ; cf. chapitre 1, section 1.1.2.) ne soit pas respectée : les énoncés qui permettent de qualifier le récit d'aventure sont trop éloignés de P3 pour pouvoir être reliés entre eux.

Dans notre corpus, une élève, de CE2 de surcroît, a réussi une « pirouette » remarquable : n'ayant pas pris en compte les critères d'enchâssement de P3 dans l'ensemble de son texte (planification à rebours échouée), elle a inséré P3 dans un contexte extérieur à son récit, à travers une phrase de dialogue émanant de la télévision, permettant ainsi à son texte de rester malgré tout cohérent, comme on le voit dans (126).


Dans les bois Lucie y habitait. Elle habitait dans cette maison depuis longtemps. La maison était très ancienne. Un beau jour elle alla chercher des pommes au marché. Elle regarda les pomme et dis« trop mur ! pas assez mur ! >> Puis fait tomber les pommes. Le marchand qui était très vieux n'entendait rien. Mais là il entandi les pommes tomber. Il se retourna en entendant ce grand br

Un que Lucie ne trouva pas sa pomme parfaite elle restra chez elle. Elle allumat la télé et entendu Depuis cette aventure, les enfants ne sortent plus la nuit.

Dans les bois Lucie y habitait. *Elle habitait dans cette maison depuis longtemps.* La maison était très ancienne. Un beau jour, elle alla chercher des pommes au marché. Elle regarda les pommes et dit : « Trop mûr ! Pas assez mûr ! ». Puis fait tomber les pommes. Le marchand qui était très vieux n'entendait rien. Mais là, il entendit les pommes tomber. *Il se retourna en entendant ce grand bruit.* Vu que Lucie ne trouva pas la pomme parfaite, elle rentra chez elle. Elle alluma la télé et entendit : « *Depuis cette aventure, les enfants ne sortent plus la nuit* ».

Figure 127 : Texte de Naïla, CE2 P

Notons par ailleurs dans ce texte que tous les référents de P1 et P2 sont correctement traités.

CHAPITRE 13 : Bilan des résultats et comparaison

Afin de répondre à notre question de recherche, à savoir comment les élèves de 9 à 11 ans introduisent et maintiennent les référents dans les productions écrites narratives (cf. chapitre 6), nous avons proposé à des élèves du CE2 au CM2 une tâche-problème qui consiste à raconter une histoire dans laquelle ils doivent insérer, séparément et dans l'ordre indiqué, les trois phrases suivantes :

- P1 : *Elle habitait dans cette maison depuis longtemps.*
- P2 : *Il se retourna en entendant ce grand bruit.*
- P3 : *Depuis cette aventure, les enfants ne sortent plus la nuit.*

Dans les chapitres 8 à 12, nous avons procédé à l'exploitation de l'annotation systématique des textes dans le but de mettre au jour les moyens linguistiques utilisés par les élèves de différents niveaux scolaires pour introduire des référents dans un discours, les gérer et les maintenir tout en construisant un récit. Nous avons donc analysé chacune des occurrences de la consigne qui nécessitait l'introduction d'un référent :

- les pronoms personnels clitiques de troisième personne *elle* et *il* contenus respectivement dans P1 et P2
- les SN démonstratifs *cette maison* et *ce grand bruit* contenus également dans P1 et P2
- les éléments de la dernière phrase (P3) : le SN défini *les enfants*, le SN défini *la nuit* et l'anaphore résomptive *depuis cette aventure*

Nous proposons dans ce chapitre de faire un état des lieux des principaux résultats obtenus : d'une part, en résumant nos analyses effectuées dans les chapitres précédents et d'autre part, en comparant la manière dont les élèves introduisent et maintiennent des référents de nature différentes dans le discours.

À cette fin, nous choisissons de procéder par un aller-retour entre nos questions de recherche et les résultats obtenus. Aussi, en 13.1., nous nous demandons quels sont les procédés d'introduction d'un nouveau référent dans un récit utilisés dans les textes étudiés. En 13.2., nous nous interrogeons sur les procédés de coréférence employés dans les textes de notre corpus. En 13.3., nous questionnons la présence ou non d'une expression référentielle

visant les référents des entités présentes dans les phrases de la consigne, afin de déterminer lesquelles sont les plus difficiles à gérer.

13.1. Quels sont les procédés d'introduction d'un nouveau référent dans un récit utilisés dans les textes étudiés ?

Les différentes études qui se sont intéressées à l'introduction des référents dans un discours chez les adultes, et plus particulièrement dans un texte narratif comme c'est le cas dans notre recherche, montrent que le SN indéfini est la forme d'introduction privilégiée. En effet, dans un genre narratif tel que le roman, il s'avère que les premiers maillons des chaînes de référence sont principalement des SN indéfinis (cf. les nombreux travaux que nous avons étudiés au chapitre 2).

Les recherches qui ont étudié l'introduction des référents chez les enfants aboutissent à la même conclusion : le SN indéfini est également la forme privilégiée de l'introduction d'un référent (cf. chapitre 3). Toutefois, l'introduction d'un référent par un SN indéfini se réalise progressivement et tardivement chez les enfants. Les travaux d'Hickman (1980 ; 1984 ; 1987 ; mais aussi 2003 ; 2004) notamment, ont mis en évidence le fait que plus les enfants sont jeunes, moins ils recourent à un SN indéfini pour introduire un nouveau référent.

Nous allons voir à travers l'étude de l'introduction des référents des occurrences de la consigne que si les travaux précédemment cités sont en accord avec nos résultats dans une certaine mesure, ce n'est pas pourtant pas toujours le cas.

13.1.1. Introduction des personnages : les référents des pronoms personnels clitiques de troisième personne *elle* et *il*, le SN défini *les enfants*

Introduction des référents des pronoms personnels **elle** *et* **il**

Dans P1 comme dans P2, les pronoms personnels *elle* et *il* nécessitent au préalable l'introduction d'un référent car l'emploi d'un pronom implique un degré d'accessibilité fort de l'entité à laquelle il réfère (cf. chapitre 1, section 1.2.2.). En effet, les pronoms personnels

ne comportent pas de tête lexicale. Ils ne portent donc pas en eux-mêmes d'indication sur la catégorie de leur référent, mais cela ne les empêche pas de renvoyer à des particuliers et donc de référer de manière définie (Charolles, 2002 ; cf. chapitre 1, section 1.4.2.2.). Toutefois, comme le pronom personnel de troisième personne est marqué en genre, son emploi engendre la présence d'une source : « la production et la compréhension d'un pronom marqué en genre suppose en effet que le locuteur et l'interprétant ont non seulement accès à la catégorie sous laquelle est conceptualisé le référent mais au terme désignant cette catégorie » (Charolles, 2002 : 185). Le pronom personnel clitique de troisième personne n'étant pas suffisamment autonome, il nécessite que le lecteur sache déjà à quelle entité il réfère pour pouvoir l'interpréter : le référent du pronom personnel doit donc être introduit au préalable dans le discours.

Si le SN défini *les enfants* implique qu'il s'agit de personnages humains, les pronoms personnels contenus dans P1 et P2 n'indiquent pas la nature des occurrences : ils pourraient tout aussi bien référer à des humains qu'à tout autre type d'entité animée, telles que des animaux ou des objets dotés de vie, comme on le voit régulièrement dans la littérature de jeunesse notamment. Pourtant, dans notre corpus, tous les élèves ont choisi d'attribuer un référent humain aux pronoms personnels de P1 et P2, à l'exception de deux textes dans lesquels le référent du pronom personnel *il* est l'un un dinosaure, l'autre un chien. La prédominance de l'attribution du caractère humain peut s'expliquer par la consigne qui oriente ce choix : le SN défini *les enfants* peut suggérer qu'il s'agit des personnages féminin et masculin de P1 et P2.

Que ce soit pour introduire le référent du pronom personnel *elle* ou bien celui du pronom personnel *il*, le SN indéfini est la forme de la première mention la plus fréquemment employée dans notre corpus, conformément à ce qui est attendu : 52% dans l'ensemble du corpus pour le pronom personnel *elle* et 31% pour le pronom personnel *il*. Pour l'un comme pour l'autre, la tête du syntagme nominal indéfini a le plus souvent la forme de ce que Charolles appelle des *dénominations passe-partout* telles que *une fille, une femme, un homme, un garçon.* Dans les deux cas, de manière générale, la fréquence d'emploi du SN indéfini augmente avec le niveau scolaire des élèves. Nos résultats nous conduisent donc à la même conclusion que les travaux d'Hickman concernant les enfants entre 6 et 10 ans (1980 ; 1984 ; 1987).

Le taux d'échec constaté entre l'introduction du pronom personnel *elle* et celle du pronom personnel *il* n'est pas le même. Cette différence s'explique par le fait que pour l'introduction du référent du pronom personnel *il,* le deuxième type de forme le plus employé (22%) est le SN possessif, notamment chez les élèves les plus jeunes. Les élèves sont donc nombreux à établir un lien entre les personnages. Cela peut être dû au fait que l'élève est amené à créer un lien entre les personnages féminin de P1 et masculin de P2 afin de garder la cohérence globale du texte, car il semble plus facile d'amener un nouveau personnage grâce à un lien de parenté (ou autre) avec un autre personnage, d'où le recours au SN possessif, plutôt que d'introduire un personnage sans lien aucun avec les autres.

Pour introduire les référents des pronoms personnels, on constate également l'emploi de noms propres, dans des proportions à peu près égales : 16% pour le référent du pronom personnel *elle* et 14% pour celui du pronom personnel *il.* Bien que l'emploi d'un nom propre est tout aussi approprié pour introduire un référent (cf. Landragin & *al.*, 2015), cette stratégie n'est pas la principale chez les élèves. Dans ces travaux sur la gestion de la référence par des élèves, Charolles (1988c ; cf. chapitre 3, section 3.1.2.3.) arrivait également à la conclusion que la dénomination à l'aide d'un nom propre est une stratégie secondaire chez les élèves.

L'étude des fonctions syntaxiques de la première mention des référents des pronoms personnels *elle* et *il* met en évidence le fait que la première mention du référent du pronom personnel *elle* est aussi bien placée en position de sujet syntaxique que dans une construction présentative telle que *Il était une fois*, alors que la première mention du référent du pronom personnel *il* est principalement en position de sujet syntaxique. Le fait que le personnage féminin est le premier à introduire dans l'histoire le place dans une position plus propice à l'emploi de constructions présentatives. La position de sujet syntaxique, quant à elle, témoigne de la saillance de ces personnages par rapport aux autres (les éléments moins saillants tels que les entités non animées sont préférentiellement placés en position de complément, comme nous le verrons plus loin) et de leur rôle de sujet thématique dans l'énoncé. C'est la stratégie du sujet thématique (Karmiloff-Smith, cf. chapitre 3, section 3.1.).

Il semble donc que les procédés utilisés pour introduire un nouveau référent dans un texte, quand celui-ci est une entité humaine, sont en accord avec les travaux antérieurs : la forme privilégiée d'introduction est le SN indéfini en position de sujet syntaxique.

Introduction du référent du SN défini **les enfants**

Toutefois, ce n'est pas le cas lorsqu'il s'agit d'introduire un référent pluriel. Du fait des particularités de P3, la première mention du référent du SN défini *les enfants* est de différentes natures :

- Le groupe nominal *les enfants* est constitué par les personnages de P1 et P2, c'est-à-dire les référents des pronoms personnels *elle* et *il*
- Le groupe nominal *les enfants* est le référent des personnages différents de ceux de P1 et P2
- Le groupe nominal *les enfants* réfère à toute une catégorie d'enfants

Notre étude montre que la forme linguistique la plus fréquente pour référer au SN défini *les enfants* est le SN numéral (24%). Il s'agit de formes telles que *trois enfants*, *cinq enfants*... On trouve également, dans des proportions moindres, des SN définis (12%) qui peuvent référer à un groupe d'enfants identifiés dans le texte, ou bien à une catégorie plus générale comme les enfants d'un village par exemple. Si dans les deux cas (SN numéral et SN défini) il s'agit de référer à un groupe, l'emploi de l'un ou de l'autre indique toutefois une différence. Le SN numéral ici peut fonctionner à la manière des dénominations passe-partout : le groupe d'enfants est uniquement identifié par le nombre d'individu qui le compose. Le SN défini quant à lui, pointe un groupe d'individu identifié par des caractéristiques plus « fines » que l'élève doit introduire dans le co-texte afin que l'identification du référent réussisse.

L'étude des fonctions syntaxiques du référent du SN défini *les enfants* montre que la première mention du référent est principalement employée soit en position de sujet syntaxique (31%), soit en position de complément de verbe (24%). Le choix de l'une ou l'autre des fonctions peut dépendre de la nature du référent (catégorie d'individus ou regroupement de personnages).

13.1.2. Introduction des entités non humaines, non animées : les SN démonstratifs *cette maison* et *ce grand bruit,* l'indication temporelle *la nuit*

Introduction des référents des SN démonstratifs **cette maison** *et* **ce grand bruit**

Les SN démonstratifs *cette maison* et *ce grand bruit* nécessitent un traitement particulier d'un point de vue sémantique. En effet, les noms têtes *maison* et *bruit* contenus

dans les SN démonstratifs de P1 et P2 évoquent une réalité qui ouvre tout un champ notionnel dans lequel les élèves peuvent venir chercher un ou plusieurs terme(s) différent(s) du nom tête. Ce terme peut alors leur servir de source pour les référents des SN démonstratifs à traiter, voire impliquer une inférence qui permet de comprendre la référence.

Le SN indéfini est la forme la plus fréquemment employée pour la première mention du référent du SN démonstratif *cette maison* (41%). En outre, mis à part quelques rares cas, le nom tête du SN indéfini utilisé, *maison,* est le même que celui du SN démonstratif. Dans certains cas tout de même, on retrouve des noms têtes tels que *château, manoir, cabane.* De surcroît, certains textes ne donnent aucune source linguistique au SN démonstratif *cette maison* mais utilisent la cooccurrence du verbe *habiter* et du SN *maison* pour permettre une inférence.

Cependant, l'analyse de l'introduction du référent du SN démonstratif *ce grand bruit* mène à une autre conclusion. La forme de la première mention du référent du SN démonstratif *ce grand bruit* la plus employée est une phrase entière qui explique la provenance du bruit (29% des textes du corpus). Le recours au SN indéfini n'est que la deuxième forme la plus fréquente (23%). On retrouve également quelques onomatopées (9%) dans tous les niveaux de classe.

L'étude des fonctions syntaxiques de la première mention du référent des SN démonstratifs montre que la position de complément (de verbe ou de phrase) est la plus fréquente. Cela peut s'expliquer par le fait que les SN démonstratifs contenus dans P1 et P2 sont eux-mêmes en position de complément (de verbe ou de phrase). Les élèves ont donc tendance à garder les mêmes fonctions pour la première mention, d'autant plus qu'elles sont toutes deux en compétition avec le référent des pronoms personnels, contenus dans les mêmes phrases. C'est la stratégie de non-changement de rôle (Kail & Léveillé, 1977, cf. chapitre 3, section 3.1.).

Introduction du référent du SN défini la nuit

Pour introduire le référent du SN défini *la nuit*, le SN défini est la forme la plus utilisée (30%). Le déterminant défini s'impose ici dans la mesure où le SN défini *la nuit* revêt un caractère générique par sa position dans la phrase fermoir P3 (cf. chapitre12).

On trouve dans une moindre mesure des SN indéfinis (12%) mais leur emploi sort de son cadre habituel. En effet, les SN indéfinis sont habituellement employés en première mention car ils sont autonomes référentiellement parlant (Charolles, 2002 ; cf. chapitre 1, section 1.4.2.1.) : ils ne demandent pas à ce que l'interlocuteur sache déjà de quoi il va être question dans le propos. En cela, ils permettent généralement de fixer le référent et non de le désigner (comme c'est le cas des SN définis et démonstratifs). Dans le cas de l'occurrence *la nuit,* employée dans un sens générique, le SN indéfini *une nuit* permet justement de désigner une nuit spécifique, de la mettre en relief par rapport à toutes les autres nuits. La valeur du SN indéfini employé dans notre cas se rapproche donc de la valeur ordinaire des SN définis et démonstratifs.

Enfin, on trouve également des phrases entières pour référer au SN défini *la nuit*, nécessitant ou non une inférence.

En ce qui concerne la position syntaxique de la première mention du référent, c'est majoritairement la même que celle occupée par le SN défini *la nuit* de P3, c'est-à-dire complément de phrase.

Ainsi, si les travaux de recherche ont révélé que le SN indéfini était la forme privilégiée pour introduire un nouveau référent animé (humain ou non) dans un discours, que ce soit chez les adultes ou les enfants, nos travaux ont montré que la forme de la première mention d'un référent dépend de sa nature (personnage, évènement, indication spatio-temporelle).

13.2. Quels sont les procédés de coréférence employés dans les textes étudiés ?

Nous l'avons déjà vu, maintenir la référence dans un discours nécessite l'emploi d'expressions coréférentielles, dont le nombre et la nature peuvent varier. Lorsqu'il y a uniquement deux mentions d'un référent, on parle d'anaphore (Reichler-Béguelin, 1988 ; cf. chapitre 2, section 2.1.). Mais à partir de trois mentions, les expressions coréférentielles forment une chaîne plus ou moins longue appelée chaîne de référence (CDR) ((Schnedecker, 1997 ; cf. chapitre 2, section 2.2.).

Les études qui se sont intéressées au maintien de la référence dans des textes narratifs de scripteurs experts montrent que les formes de reprise du référent diffèrent en fonction de la nature plus ou moins saillante du référent en question. La chaîne de référence d'un topique « continu » (soit un référent saillant) devrait être majoritairement composée de pronoms personnels. Les chaînes de référence d'un topique discontinu (ou référent non saillant) devraient, quant à elles, être plus courtes et instancier davantage de SN que de pronoms car elles entrent en compétition avec d'autres chaînes (Givón, 1983 ; Schnedecker et Landragin, 2014 ; cf. chapitre 2, section 2.2.4.).

Les mêmes conclusions apparaissent dans les stratégies des apprentis scripteurs, mais là encore, la maîtrise du maintien de la référence se réalise de manière progressive. Les travaux de Karmiloff-Smith (1980, 1981) en particulier, montrent que les enfants de 6 ans et au-delà réservent exclusivement l'emploi du pronom personnel placé en début de phrase au sujet thématique, qui est en général le personnage principal de leur histoire, c'est-à-dire le référent saillant. Les autres entités du récit, moins saillantes, sont dénommées à l'aide de SN définis. Les enfants plus âgés (8/9 ans) emploient également le pronom en position initiale dans la phrase pour référer au sujet thématique, mais pas de manière exclusive. Ils peuvent aussi placer des personnages secondaires en position initiale, qu'ils décrivent alors à l'aide d'expressions définies. Jisa (2000) a montré que chez les enfants, l'article défini pouvait également être employé pour réintroduire un référent, c'est-à-dire placer à nouveau en position sujet un référent déjà introduit dans le discours. Ou en d'autres termes, réactiver le référent devenu peu accessible en mémoire.

Nous allons voir à travers l'étude du maintien de la référence que le nombre moyen d'expressions coréférentielles et leur nature est fonction des caractéristiques du référent.

13.2.1. Anaphores ou CDR ? Nombre moyen de mentions

L'étude du nombre moyen de mentions du référent pour chaque occurrence montre des différences de traitement entre les occurrences (cf. tableau 51). Les référents des personnages humains sont dans l'ensemble maintenus plus longtemps dans les textes que les référents des entités non humaines non animées.

Occurrence	Nombre moyen de mentions du référent	Nombre de mentions dans la CDR la plus longue
Pronom personnel *elle*	4,8	21
Pronom personnel *il*	5,0	26
SN défini *les enfants*	3,3	19
SN démonstratif *cette maison*	2,1	9
SN démonstratif *ce grand bruit*	1,8	11
SN défini *la nuit*	1,5	4

Tableau 51 : Étude de la longueur moyenne des CDR et nombre de mentions dans la CDR la plus longue, pour chaque occurrence

***Longueur moyenne des CDR des pronoms personnels* elle *et* il**

Les personnages de P1 et P2 font l'objet du plus grand nombre de mentions : environ 5 mentions en moyenne. Ce sont donc les occurrences qui sont maintenues le plus longtemps dans les textes, à l'aide de véritables CDR qui permettent une réelle construction des personnages et un traitement global. Les référents des pronoms personnels deviennent les personnages centraux du récit. Cela paraît naturel dans la mesure où les pronoms personnels de P1 et P2 sont placés en position initiale dans les phrases imposées, occupent la fonction de sujet syntaxique et sujet thématique, et sont caractérisés par leur saillance du fait de leur nature humaine. Ils orientent donc la construction du récit pour que leurs référents deviennent également le sujet thématique de l'énoncé dans sa totalité.

Nous remarquons que la longueur moyenne varie en fonction du niveau de classe : au CE2, la longueur moyenne de la CDR des pronoms personnels est d'environ 4 maillons alors qu'elle est d'environ 5 maillons au CM1 et au CM2. Cet écart peut s'expliquer par l'accroissement des capacités cognitives et de gestion de la cohésion textuelle avec l'âge (*knowledge telling*/*knowledge transforming*, Bereiter & Scardamalia, 1987 ; cf. chapitre 3, partie 3.1.4.), plus le niveau de classe et l'âge des élèves augmente, plus ils construisent une véritable CDR, de plus en plus longue.

***Longueur moyenne de la CDR incluant le SN défini* les enfants**

En ce qui concerne le SN défini *les enfants*, on constate des CDR globalement plus courtes que pour le maintien aux référents des pronoms de P1 et P2. La longueur moyenne des

CDR incluant le SN défini *les enfants* est d'environ 3 maillons : il s'agit là de CDR minimales. Ce résultat s'explique par le caractère conclusif de P3 qui suggère que le référent du SN défini *les enfants* ne soit pas maintenu très longtemps.

On constate également que la longueur moyenne des CDR augmente avec le niveau scolaire des élèves : elle passe d'environ 2 maillons au CE2 à 3 maillons aux CM1 et CM2. La différence constatée peut s'expliquer par les propriétés de P3 qui en font une phrase difficile à traiter en termes de cohérence textuelle. Cette complexité est plus difficile à gérer pour les élèves les plus jeunes, ils prennent donc moins le temps de « dérouler » les CDR des entités qu'elle contient, d'autant plus que l'occurrence *les enfants* comprend déjà en soi plus de difficultés que les « simples » pronoms personnels de P1 et P2.

***Longueur moyenne des CDR incluant les SN démonstratifs* cette aventure *et* ce grand bruit**

Les référents des SN démonstratifs *cette maison* et *ce grand bruit* sont maintenus de manière anaphorique : le nombre moyen de mentions est d'environ 2. Les SN démonstratifs représentent des occurrences plutôt secondaires dans l'histoire, notamment par leur caractère non humain et non animé, en opposition, nous l'avons vu, avec les personnages humains des phrases imposées. Mais également par les positions syntaxiques de complément qu'ils occupent dans P1 et P2.

En outre, le SN démonstratif *ce grand bruit* est plus complexe à traiter du fait de la difficulté qu'il peut y avoir à parler d'un bruit, d'autant plus lorsqu'il s'agit d'un *grand* bruit. La longueur moyenne de sa CDR est donc moins grande (1,8 maillons) que celle des CDR incluant le SN démonstratif *cette maison* (2,1 maillons). Toutefois, les CDR les plus longues peuvent comporter jusqu'à 8 mentions pour celle de l'occurrence *maison* et jusqu'à 10 mentions pour celle de l'occurrence *grand bruit*.

***Longueur moyenne de la CDR incluant le SN défini* la nuit**

Le SN défini *la nuit* est l'occurrence qui fait l'objet du plus petit nombre de mentions : 1,5 mentions. Il est donc plutôt repris de manière anaphorique (2 mentions dans le texte). Le statut d'indication temporelle de l'occurrence *nuit* en fait une occurrence qu'il n'est pas nécessaire de répéter plus de 2 fois dans le discours : la mention contenue dans P3 et son

introduction peuvent suffire, à moins bien entendu de raconter une histoire dans laquelle *la nuit* est le sujet thématique.

Après avoir donné une vue globale de la manière dont est maintenue la référence à chaque entité, nous donnons une vue d'ensemble de la composition des CDR des occurrences à traiter.

13.2.2. Compositions des CDR des entités de la consigne

***Maintien de la référence aux entités humaines : les pronoms personnels* elle *et* il*, le SN défini* les enfants**

Que ce soit pour le pronom personnel *elle*, le pronom personnel *il* ou le SN défini les *enfants*, le SN pronominal reste la forme privilégiée de la reprise, le plus souvent en position de sujet syntaxique, excepté lorsque la chaîne de référence du personnage féminin ou masculin entretient des relations d'entrecroisement avec des chaînes référant à d'autres entités.

Les formes des mentions des référents des pronoms peuvent également varier (SN définis, noms propres notamment), dans une moindre mesure, en fonction des nécessités de réinstanciation du personnage, que ce soit pour des raisons de saillance (distance minimale) ou de changement de personnage dans le focus d'attention dû à l'entrecroisement de plusieurs chaînes de référence appartenant à diverses entités.

L'usage dominant du pronom, à partir de la deuxième mention, n'est pas étonnant puisqu'il était déjà admis que cette forme est la plus couramment employée dans les chaînes de référence des textes narratifs (cf. introduction de ce chapitre). En effet, les personnages féminin de P1 et masculin de P2, par leur position en début de proposition, représentent un référent saillant. Comme ils ont été introduits précédemment à P1 et P2, leur position saillante présente un haut degré d'accessibilité. Le pronom est donc la forme la plus indiquée pour la reprise, chez les scripteurs experts comme chez les apprentis scripteurs (Givon, 1983, 1989 ; Schnedecker & Landragin, 2014 ; Karmiloff-Smith, 1980 ; 1981). Pour les mêmes raisons, on retrouve les mentions du référent principalement en position de sujet syntaxique, confirmant ainsi la stratégie du sujet thématique employée par les élèves (Karmiloff-Smith, 1980, 1981 ; et plus récemment Millogo, 2008 ; cf. chapitre 3, section 3.1.2.3.) : le pronom est utilisé

comme forme de reprise du sujet thématique, placé en position initiale dans la phrase à travers la position de sujet syntaxique.

Maintien de la référence aux entités non humaines, non animées : SN démonstratifs cette maison, ce grand bruit et SN défini la nuit

Dans les CDR incluant le SN démonstratif *cette maison*, comme dans celles du SN défini *la nuit*, on constate une prédominance de SN définis et de pronoms comme formes de reprise du référent, principalement en position de complément (de phrase ou de verbe). Conformément aux travaux de Karmiloff-Smith notamment (cf. supra), ce résultat est en accord avec la nature moins saillante de *maison* et *nuit*.

En revanche, dans les CDR incluant le SN démonstratif *ce grand bruit*, les SN définis, démonstratifs, onomatopées et phrases explicatives sont aussi fréquents les uns que les autres, certainement à cause du fait que le nom *bruit* est un nom d'évènement qui suppose des formes de reprise variées. Toutefois, la comparaison par niveaux montre que les formes telles que SN définis et démonstratifs dans les CDR ont plutôt tendance à apparaître chez les élèves les plus âgés. L'étude des fonctions syntaxiques des différentes mentions du référent du SN démonstratif *ce grand bruit* montre également qu'il n'y a pas de prédominance de l'une ou l'autre des fonctions : on retrouve indifféremment des formes en position de sujet ou de complément de verbe.

Après avoir étudié les procédés d'introduction et de maintien de la référence à des entités de natures différentes, nous proposons une vue générale de la gestion de la référence à ces entités en nous interrogeant sur le taux d'échec de l'introduction des référents en fonction de chaque occurrence.

13.3. Les expressions linguistiques de la consigne font-elles toutes l'objet d'au moins une mention dans le texte, mise à part celle de la phrase imposée ?

Nous l'avons vu, la consigne impose au scripteur d'introduire dans son récit des expressions référentielles visant les référents des expressions présentes dans les phrases qu'il

doit insérer. Comme nous l'avons déjà souligné plus haut, les entités des phrases imposées sont de natures très diverses :

- des personnages animés (*elle*, *il*) et humains (*les enfants*)
- une localisation spatiale (*dans cette maison*)
- un évènement (*ce grand bruit*)
- une information temporelle (*la nuit*)

Nous l'avons dit, sur le plan sémantique, toutes ces entités n'ont pas le même degré de saillance. En effet, les caractéristiques des entités portées par les mots ont une influence sur leur saillance (Lyons, 1980 ; Pattabhiraman, 1993 ; cf. chapitre 1, section 1.2.4.). Le caractère « humain » ou « animé » d'une entité augmente sa saillance, par rapport à une entité « non animée » ou un animal (Lyons, 1980).

En outre, toutes les phrases imposées de la consigne ne présentent pas les mêmes caractéristiques. P1 regroupe une entité animée et une localisation spatiale, P2 regroupe une entité animée et un évènement. Dans chacune des phrases imposées, il y a donc deux entités de natures différentes à introduire. Quant à P3, non seulement elle contient des personnages humains ainsi qu'une information temporelle, mais elle comporte également l'anaphore résomptive *depuis cette aventure*, qui lui donne un caractère conclusif de phrase fermoir (Marandin, 1987).

Tous ces éléments sont à prendre en compte dans le récit mais nos résultats montrent que toutes les expressions linguistiques ne font pas nécessairement l'objet d'une mention dans les textes, dans des proportions différentes, comme on peut le voir dans le tableau 52.

Taux d'échec d'introduction des référents dans le discours								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
Elle	7	11%	4	7%	3	5%	14	8%
Il	6	9%	10	18%	4	6%	20	11%
Cette maison	17	27%	18	32%	13	20%	48	26%
Ce grand bruit	13	19%	24	34%	19	27%	56	27%
Les enfants	21	33%	18	32%	28	43%	67	36%
La nuit	26	41%	29	52%	31	48%	86	46%

Tableau 52 : Taux d'échec[57] de l'introduction des référents dans le discours

La lecture du tableau montre que :

- le référent du pronom personnel *elle* est presque toujours introduit, bien que les élèves de CE2 soient les plus nombreux à ne le pas faire
- le référent du pronom personnel *il* est assez bien introduit également, sauf pour les CM1
- le référent du SN défini *les enfants* est souvent non introduit, particulièrement chez les CM2
- la référence au SN défini *la nuit* est le plus souvent omis

Nous allons à présent analyser plus précisément chacun des constituants, en débutant par les occurrences qui ont le plus souvent été omises.

13.3.1. Les constituants de la phrase *fermoir* P3

Les expressions linguistiques qui ne font l'objet d'aucune mention dans le texte produit sont surtout les expressions présentes dans la phrase fermoir P3 : le SN défini *les enfants* et le SN défini *la nuit*.

L'indication temporelle *la nuit* est le constituant qui a le plus fort taux d'échec d'introduction : 46%. Plusieurs raisons peuvent l'expliquer. Premièrement, le caractère de

[57] Dans le tableau 52, nous avons recensé, pour chaque occurrence, tous les cas où la phrase imposée est la seule mention du référent dans le texte des élèves. Les pourcentages sont exprimés en fonction du nombre d'introduction « réussies » dans chaque occurrence. Ils ne peuvent donc être additionnés.

circonstant temporel fait du SN défini *la nuit* une entité secondaire, « d'arrière-plan », par rapport aux autres constituants de la consigne. Il représente aussi une information substituable par sa position syntaxique de complément de phrase. Deuxièmement, le SN défini *la nuit* apparaît en position finale, non seulement de P3, mais également de toutes les phrases de la consigne : c'est le dernier constituant que l'on identifie à la lecture des phrases imposées. Enfin, il apparaît également dans une phrase aux caractéristiques propres : P3 n'a pas la même valeur que P1 ou P2, à cause de son caractère conclusif. C'est la phrase qui demande le plus de ressources cognitives : par exemple, si on la découvre alors que le texte est déjà rédigé, il est plus difficile de l'ajouter.

Le taux d'échec dans l'introduction du référent du SN défini *les enfants* est de 36%, rendant cette occurrence la deuxième la moins bien gérée. Cela peut paraître surprenant dans la mesure où le référent à introduire est non seulement un référent humain, donc plus saillant, mais il est aussi en position de sujet syntaxique, faisant ainsi de lui le sujet thématique. Toutes les conditions semblent réunies pour que, à l'instar des personnages masculins et féminins, les cas de non prise en compte du référent du SN défini *les enfants* soient faibles. Pourtant, ce n'est pas le cas. On peut peut-être expliquer ce résultat par différents facteurs.

D'abord, le fait que la consigne oriente le référent du SN défini *les enfants* vers le groupe constitué par les personnages de P1 et P2. En effet, si la consigne reste libre, elle offre toutefois la possibilité de faire des personnages les enfants dont il est question en P3, puisque les pronoms personnels *elle* et *il* sont saillants, de même que le SN défini *les enfants*, alors que les autres constituants des phrases imposées ne le sont pas. La réunion de deux entités introduites séparément (les référents des pronoms personnels *elle* et *il*) au sein d'un groupe pluriel (*les enfants*) n'est pas une opération facile pour les apprentis scripteurs.

Ensuite, de même que pour l'indication temporelle *la nuit*, le SN défini *les enfants* apparaît dans P3. Cela signifie non seulement que c'est une des dernières entités que les élèves identifient mais également, que ses caractéristiques, nous l'avons vu, mobilisent beaucoup de ressources cognitives.

13.3.2. Les référents des SN démonstratifs *cette maison* et *ce grand bruit*

La non prise en compte des SN démonstratifs *cette maison* et/ou *ce grand bruit* est un peu moins fréquente. En effet, 26% des textes ne comportent pas d'expression linguistique en relation de coréférence avec le SN démonstratif *cette maison* et 27% avec le SN démonstratif *ce grand bruit*.

Le léger écart constaté s'expliquer par le fait que le SN démonstratif *ce grand bruit* est plus complexe à traiter. D'une part, le nom *bruit* est un nom d'évènement qui réfère à une situation plutôt qu'à une entité concrète. Le caractère plus central pour le récit, en opposition avec le caractère plus local du SN démonstratif *cette maison*, nécessite donc de construire une « scène » qui produit le bruit. D'autre part, la présence de l'expansion *grand* contraint la réalisation du SN démonstratif *ce grand bruit* : un grand bruit n'est pas n'importe quelle sorte de bruit.

Par leur nature d'entité non humaine et non animée, les deux SN démonstratifs sont des entités moins saillantes que les personnages de P1 et P2 par exemple, d'autant plus qu'ils sont placés en position de complément (de verbe ou de phrase) alors que les personnages sont en position de sujet syntaxique. En revanche, les SN démonstratifs étant des arguments des verbes *habiter* et *entendre*, leur caractère moins périphérique, par rapport aux constituants de P3, leur permet d'être relativement mieux gérés par les élèves.

13.3.3. Les référents humains des pronoms personnels clitiques de troisième personne *elle* et *il*

Les expressions linguistiques qui ne font l'objet d'aucune mention dans les textes produits, en dehors de celle contenues dans les phrases imposées, sont, de manière tout à fait marginale, les expressions qui réfèrent aux pronoms personnels *elle* de P1 et *il* de P2. En effet, 8% seulement des textes du corpus ne comportent pas d'expression linguistique en relation de coréférence avec le pronom personnel *elle*. Il en va à peu près de même pour le pronom personnel *il* : 11% des textes.

Le personnage féminin semble légèrement plus facile à traiter que le personnage masculin. Cela peut peut-être s'expliquer par la position initiale de P1 : le pronom personnel *elle* est la première entité que les élèves repèrent alors que le pronom personnel *il* est la troisième. L'introduction du référent du personnage masculin est donc moins facile à réaliser que celle du personnage féminin puisque des ressources cognitives sont déjà mobilisées pour d'autres opérations : construire un récit, introduire le référent du pronom sujet *elle*, parler d'une maison.

Les textes dans lesquels il n'y a aucune mention du référent mise à part celle contenue dans la phrase imposée présentent un caractère de non cohérence. En effet, les personnages masculin et féminin de la consigne apparaissent dans P1 et P2 sous la forme de pronoms personnels clitiques de troisième personne. L'interprétation d'un pronom personnel clitique de troisième personne exige que le lecteur sache à quelle entité il réfère ou, en d'autres termes, que le référent du pronom personnel ait été introduit au préalable dans le discours. De surcroît, les pronoms personnels de P1 et P2 occupent la fonction de sujet syntaxique de la phrase, ce qui en fait également le sujet thématique. Si le texte ne tient pas compte d'une de ces occurrences, la cohérence textuelle n'est pas assurée.

13.3.4. Nombre d'occurrences introduites dans les textes

Nous avons montré que toutes les expressions linguistiques de la consigne ne faisaient pas nécessairement l'objet d'une mention dans le texte. Sur les six expressions référentielles contenues dans les phrases imposées de la consigne, certains textes n'en intègrent que quelques-unes. Le tableau 53 montre combien d'expressions référentielles sur les six intégrées dans le texte.

Nombre d'occurrences intégrées dans les textes								
	CE2		**CM1**		**CM2**		**Total**	
	Nbr	**%**	**Nbr**	**%**	**Nbr**	**%**	**Nbr**	**%**
1 sur 6	**0**	0%	2	4%	1	2%	**3**	2%
2 sur 6	2	3%	4	7%	1	2%	**7**	4%
3 sur 6	9	14%	13	23%	10	15%	32	17%
4 sur 6	19	**30%**	9	16%	19	29%	47	25%
5 sur 6	17	27%	20	**36%**	21	**32%**	58	**31%**
6 sur 6	17	27%	8	14%	13	20%	38	21%
Total	64	100%	56	100%	65	100%	185	100%

Tableau 53 : Nombre d'expressions référentielles intégrées dans le texte, en fonction du niveau de classe

À la lecture du tableau, on constate que quel que soit le niveau de classe, les textes dans lesquels il n'y a qu'une ou deux occurrences de la consigne qui sont prises en compte sont très peu nombreux : sur l'ensemble du corpus, il n'a que 3 textes dans lesquels une seule occurrence est introduite sur les six et 7 textes dans lesquels deux occurrences sont introduites sur les six. Remarquons qu'au CE2, tous les textes introduisent au moins une occurrence de la consigne.

Globalement, dans la majorité des textes, les sources de cinq occurrences sur les six de la consigne sont localisées : 31% des textes introduisent cinq référents. Lorsqu'on compare les différents niveaux de classe, on constate qu'au CE2, la majorité des textes intègre quatre occurrences (30%), alors que la majorité des textes de CM1 et de CM2 en intègre cinq (36% au CM1 et 32% au CM2). Il semble donc que le niveau de classe influe sur le nombre d'occurrences introduites dans les textes.

Au terme de ce chapitre, nous avons dressé un état des lieux des principaux résultats de notre recherche qui montrent que la gestion de la référence dépend des caractéristiques des référents mais également de la prise en compte ou non du récit dans sa globalité. Comme le fait remarquer Wigglesworth (1990 ; cf. chapitre 3, section 3.1.2.1.), l'aptitude à raconter un récit cohérent et structuré requiert des compétences linguistiques et cognitives considérables : l'information doit être organisée en un tout, les participants nommés et différenciés et les référents linguistiques établis pour les relations anaphoriques futures. Il nous semble que ce dernier point est important car de la bonne introduction du référent découle le bon

enchaînement des relations anaphoriques avec ses autres mentions et par là-même, la bonne cohésion de l'ensemble de l'énoncé. Cependant, introduire un référent de manière correcte n'est pas aisé. Jisa (2000) insiste sur le fait que l'introduction d'un référent nécessite des connaissances pragmatiques sophistiquées sur les éléments partagés ou non par le narrateur et le destinataire.

CONCLUSION

L'objectif de notre recherche était d'étudier l'acquisition de compétences discursives d'élèves du CE2 au CM2 en matière de cohésion textuelle, dans une perspective de progressivité. Plus précisément, nous avons étudié un des aspects de la cohésion textuelle : la continuité référentielle. Dans la première partie de la thèse, nous avons posé les éléments épistémologiques de notre recherche et montré en quoi la continuité référentielle est facteur de cohésion textuelle, elle-même facteur de cohérence textuelle. Nous avons également montré que l'acquisition des marques de la référence chez les enfants était progressive et tardive. D'un point de vue didactique, nous nous sommes intéressés à l'enseignement des marqueurs référentiels ainsi qu'à la place de la cohésion et de la cohérence textuelles dans l'évaluation et le jugement de cohérence des enseignants. Enfin, d'un point de vue méthodologique, nous avons exposé les enjeux de la constitution de corpus de textes d'élèves et les difficultés que soulève leur constitution et leur analyse.

Partant du constat que peu d'études s'intéressent à l'acquisition des marqueurs référentiels en production écrite d'un point de vue didactique et dans une perspective de progressivité, nous avons cherché à savoir comment les élèves du CE2 au CM2, soit de 9 à 11 ans, introduisent et maintiennent des référents de diverses natures dans les productions écrites narratives, en annotant systématiquement les marques anaphoriques dans un vaste corpus de textes d'élèves.

À cette fin, nous avons repris une tâche-problème élaborée par Garcia-Debanc (2012), que nous avons proposées à 210 élèves. La tâche-problème consiste en la rédaction d'une histoire dans laquelle trois phrases imposées doivent être insérées. Ces trois phrases contraignent les élèves à résoudre plusieurs problèmes d'anaphores dans la mesure où elles contiennent chacune plusieurs référents qui appellent une introduction préalable. Les référents sont de natures variées : personnages féminin et masculin (pronom *elle* et *il*), groupe de personnage (SN défini *les enfants*), mais aussi évènement (SN démonstratif *ce grand* bruit) et indications spatio-temporelle (SN démonstratif *cette maison*, SN défini *la nuit*). Ainsi, l'absence de référence explicite pour chaque entité à traiter conduit les élèves à insérer des phrases supplémentaires avant chacune d'entre elles. La résolution de l'anaphore peut donc s'effectuer de manière locale (d'une phrase à l'autre) ou globale (plusieurs phrases), à l'aide d'une ou plusieurs expressions coréférentielles (anaphore ou chaîne de référence).

Aussi, la présente conclusion s'attache à rendre compte des principaux résultats obtenus dans notre étude. Dans un second temps, nous considérons successivement les limites et les apports de notre recherche aux domaines linguistiques, psycholinguistiques et didactiques. Puis nous les mettons en perspective des prolongements qui pourront être réalisés dans l'objectif d'autres recherches sur la cohésion textuelle chez les apprentis scripteurs.

Synthèse des résultats

Pour mettre au jour les procédés employés par les élèves du CE2 au CM2 dans les textes narratifs en matière de continuité référentielle, nous avons annoté de manière systématique chacune des expressions référentielles employées tout au long du texte pour référer aux occurrences de la consigne. Nous avons ainsi pu décrire leurs formes, leurs fonctions syntaxiques ainsi que leurs nombres afin d'étudier la manière dont les référents sont introduits et maintenus dans le discours. Nous avons ensuite réalisé une étude comparée des résultats obtenus pour chacune des occurrences de la consigne. En marge de cet axe de recherche, nous avons également soulevé la question des problèmes d'interprétation référentielle, facteur de manque de cohérence textuelle.

Au niveau de l'introduction des référents, nous avons cherché à déterminer les différents degrés de maîtrise dans l'introduction des référents en fonction du niveau scolaire des élèves, en répondant aux questions suivantes : quel type d'expression référentielle les élèves emploient-ils en première mention ? Le caractère humain/non humain ou animé/non animé des référents a-t-il une influence sur le choix des formes linguistiques de la première mention ? Nous avions fait l'hypothèse que les formes linguistiques employées pour introduire un référent dépendent de la nature de l'entité en question et, dans une certaine mesure, du niveau de classe. L'analyse de l'introduction des référents a montré à la fois des éléments conformes aux travaux antérieurs et des éléments nouveaux.

Les études linguistiques indiquent que le SN indéfini (et dans une moindre mesure, le nom propre), est la forme la plus indiquée pour introduire un nouveau référent chez les scripteurs experts. Les études psycholinguistiques s'intéressant aux choix des expressions référentielles dans l'introduction de nouveaux référents chez les enfants montrent que le SN indéfini est la forme privilégiée, bien que la fréquence de son emploi soit progressive : plus les enfants sont jeunes, moins ils utilisent le SN indéfini de manière systématique. Nous retrouvons ces résultats dans nos travaux. Les référents des personnages de la consigne, c'est-à-dire les référents des pronoms personnels *elle* et *il*, sont le plus souvent introduits par un SN

indéfini. Les élèves les plus jeunes recourent moins à ce type de forme que les élèves plus âgés.

Toutefois, lorsqu'il s'agit du référent du pronom personnel masculin, qui arrive en troisième position dans la consigne, les résultats sont plus contrastés. Si le SN indéfini reste la forme la plus employée, le SN possessif est également très fréquent, notamment chez les plus jeunes. L'ordre des référents dans la consigne semble donc influer sur les expressions référentielles employées : il semble plus facile d'introduire un deuxième référent entretenant un lien avec le premier (SN possessif) que « repartir de zéro » en introduisant un deuxième référent qu'il faudra penser à relier au premier d'une manière ou d'une autre dans le texte pour que la cohérence textuelle soit respectée.

L'emploi majoritaire de SN indéfini pour introduire un nouveau référent est également constaté pour le référent du SN défini *cette maison*, avec quelques formes lexicales plus variées chez les élèves les plus âgés (*manoir, château, cabane*).

En revanche, les expressions référentielles utilisées dans l'introduction des autres référents de la consigne varient. Pour le SN défini *les enfants*, la forme la plus globalement employée est un SN numéral, tel que *trois enfants* par exemple. La marque du pluriel porté par le SN peut expliquer l'emploi d'un SN numéral au lieu d'un SN indéfini, pourtant attendu dans le cas de l'introduction d'un nouveau référent humain. Les élèves les plus âgés sont toutefois ceux qui utilisent le moins fréquemment le SN numéral.

Pour le SN démonstratif *ce grand bruit*, en plus des expressions référentielles comportant un SN (indéfini, défini, démonstratif), on trouve des formes moins ordinaires, telles que des onomatopées ou des phrases entières. La forme de la première mention du référent la plus utilisée est une phrase qui explique la provenance du bruit, comme *le tueur cassa un vase*. La prédominance de cette manière d'introduire le référent peut avoir plusieurs sources. D'abord, le SN démonstratif *ce grand bruit* décrit un évènement, une situation qui peut nécessiter la description d'une scène entière afin de comprendre la provenance du bruit. Ensuite, l'expansion *grand* qui accompagne le nom *bruit* lui confère une spécificité : il ne peut pas s'agir de n'importe quelle sorte de bruit. Le vocabulaire de l'élève étant encore en construction (voir notamment David, 2000b), il peut avoir du mal à trouver une forme lexicale appropriée pour introduire le SN sans en répéter le nom tête, comme *une explosion* par exemple. Plus les élèves sont jeunes, plus ils ont tendance à introduire le SN démonstratif *ce grand bruit* de cette manière.

Pour le SN défini *la nuit*, la forme la plus utilisée est le SN défini, très certainement en lien avec la position du SN dans la phrase *fermoir* P3 qui confère au nom *nuit* un aspect générique : les enfants ne sortent plus *la nuit*, il s'agit de toutes les nuits et non d'une en particulier.

On peut conclure premièrement que les expressions référentielles employées pour l'introduction d'un référent dépendent de sa nature. Deuxièmement, l'utilisation d'une expression référentielle appropriée pour l'introduction d'un référent se réalise progressivement chez les élèves : plus le niveau scolaire augmente, plus la fréquence d'emploi des formes linguistiques appropriées augmente.

En ce qui concerne le maintien de la référence, nous avons cherché à déterminer les différentes stratégies employées par les élèves dans les chaînes de référence en fonction du niveau de classe, en répondant aux questions suivantes : pour référer, les élèves utilisent-ils des anaphores (2 mentions) ou bien de véritables chaînes de référence (3 mentions et +) ? Quelle est la composition des chaînes de référence (longueur moyenne, type d'expressions référentielles) ? La nature des référents influe-t-elle sur la longueur de la chaîne de référence ? Nous avions fait l'hypothèse que le nombre d'expressions référentielles employées pour maintenir la référence (anaphore ou CDR) dépend du niveau de classe et que la nature des référents influe sur la longueur et la composition de la chaîne. Nous nous attendions donc à des CDR différentes en fonction des occurrences.

L'étude de la longueur des CDR montre effectivement que la nature du référent (personnage humain, évènement ou indication spatio-temporelle) influe sur le nombre de mentions. Les personnages féminin et masculin font l'objet de véritables CDR : ce sont celles qui comportent le plus grand nombre de mentions dans le texte (environ 5). Le référent du SN défini *les enfants* quant à lui, est plutôt maintenu à l'aide de CDR minimales : elles comportent en moyenne 3 mentions du référent.

Les autres référents de la consigne, à savoir les référents des SN démonstratifs *cette maison* et *ce grand bruit,* et le référent du SN défini *la nuit,* font plutôt l'objet d'anaphores : 2 mentions du référent dans l'ensemble du texte. Le caractère « secondaire » de ces référents, par rapport aux référents plus saillants que sont les personnages de la consigne, explique cette différence de traitement. En effet, par leur nature d'indication spatio-temporelle et de nom d'évènement, on ne s'attend pas à ce que ces référents soient maintenus très longtemps dans

le discours. L'étude de la longueur moyenne des CDR révèle également que quelle que soit la nature du référent, les élèves les plus jeunes ont principalement les CDR les moins longues.

L'étude de la nature des CDR, c'est-à-dire l'étude des formes linguistiques employées pour maintenir la référence, montre également des résultats en accord avec les travaux existants et des résultats inédits. Les travaux en linguistique qui s'intéressent au maintien de la référence à des entités humaines chez les scripteurs experts montrent que la forme de reprise est principalement le pronom. Chez les enfants, les travaux en psycholinguistique indiquent que le pronom est également la forme privilégiée mais pour les référents les plus saillants seulement : c'est la stratégie du sujet thématique. Les référents plus secondaires, c'est-à-dire moins saillants, sont repris à la fois par des pronoms et des SN définis.

L'étude des formes les plus fréquentes dans les CDR montre globalement des résultats en accord avec ces travaux. Les référents humains de la consigne (pronoms personnels *elle* et *il*, SN défini *les enfants*) sont placés en position de sujet thématique dans les phrases imposées et sont les entités les plus saillantes. Conformément à la stratégie du sujet thématique, la forme la plus employée pour maintenir le référent à ces entités est le pronom, quel que soit le niveau de classe des élèves. On constate tout de même des cas où les expressions référentielles varient au cours de la CDR. Il s'agit alors d'un besoin de réinstanciation du référent car il est devenu moins saillant. On trouve ce type de cas principalement chez les élèves les plus âgés. Ce résultat montre d'une part, que plus les élèves sont jeunes, moins ils varient les expressions référentielles employées dans les CDR. D'autre part, plus les élèves sont jeunes, moins ils semblent sensibles aux nécessités de réinstanciation lorsque le cas se présente, ou bien ne savent pas comment le traiter.

Dans les formes du maintien de la référence au SN démonstratif *cette maison* et au SN défini *la nuit*, les formes les plus fréquentes pour maintenir la référence sont le pronom et le SN défini, en accord avec le caractère moins saillant de ces entités.

En revanche, les formes du maintien de la référence au SN démonstratif *ce grand bruit* sont variées : les SN définis, démonstratifs, onomatopées et phrases explicatives sont aussi fréquents les uns que les autres, certainement à cause du fait que le nom *bruit* est un nom d'évènement qui suppose des formes de reprise variées. Cependant, mise à part pour la phrase explicative, on constate que les élèves les plus jeunes ont plutôt tendance à employer des formes plus « ordinaires » pour référer au SN démonstratif *ce grand bruit* (SN défini et démonstratif).

On peut conclure que le nombre d'expressions coréférentielles varie en fonction des caractéristiques du référent, quel que soit le niveau de classe. On peut dire également que la longueur moyenne des CDR augmente en fonction du niveau de classe des élèves, mais aussi du caractère plus ou moins saillant du référent.

D'autre part, la comparaison entre les taux d'introduction de chacune des occurrences de la consigne est très éclairante puisqu'elle met au jour des différences de traitement entre les diverses occurrences. En effet, l'étude des procédés de continuité référentielle employés par les élèves nous a amenées à procéder à la comparaison du taux d'introduction de chacune des occurrences en nous demandant si les expressions linguistiques de la consigne faisaient toutes l'objet d'au moins une mention dans le texte, mise à part celle de la phrase imposée. Nous avions fait l'hypothèse que la nature variée des constituants analysés et leur place dans les phrases imposées auraient une influence sur leur taux de prise en compte dans les textes. Les résultats obtenus montrent que les personnages féminin et masculin sont les plus fréquemment introduits dans le texte, suivis des SN démonstratif *cette maison* et *ce grand bruit*. Les occurrences contenues dans P3 sont celles qui ont le plus fort taux d'échec : le SN défini *la nuit* premièrement, suivi du SN défini *les enfants*. Cela peut s'expliquer principalement par le caractère conclusif de la troisième phrase imposée (P3), qui n'a pas la même valeur que les deux premières phrases imposées (P1 et P2) et qui exige de mobiliser des ressources cognitives en plus grand nombre, notamment parce qu'elle nécessite une planification à rebours du texte. La comparaison du nombre d'occurrences intégrées dans les textes, sur les six occurrences que contient la consigne, montre que plus les élèves sont jeunes, moins ils introduisent de référents. Les élèves de CE2 intègrent en moyenne quatre occurrences sur six et les élèves de CM1 et de CM2 en intègrent en moyenne cinq sur six.

En marge des procédés de continuité référentielle employés par les élèves, l'étude a également mis en lumière des problèmes d'interprétation référentielle de diverse nature. Les plus courants sont des problèmes d'ambiguïté référentielle : le texte présente plusieurs candidats à la référence et aucun élément du contexte ou du co-texte ne peut aider à la résolution. Un autre problème d'interprétation, caractéristique des apprentis scripteurs, est la non-perception des contraintes morphosyntaxiques du référent. En effet, certains élèves ne prennent pas en compte les marques de genre ou de nombre que portent les entités de la consigne. De ce fait, ils introduisent des référents aux caractéristiques non conformes avec celles des entités de la consigne, entraînant des problèmes de cohérence textuelle.

La question de la coréférence, lorsqu'elle n'est pas maîtrisée, peut également entraîner des problèmes d'interprétation référentielle. Dans certains textes, sur le plan sémantique, on peut comprendre qu'il s'agit du même référent tout au long du texte. Pourtant, sur le plan linguistique, la première mention du référent est complètement différente des autres, elle ne porte pas non plus les mêmes marques de genre et de nombre que l'entité de la phrase imposée. C'est le cas par exemple lorsque la première mention du référent est *un cadavre* et la deuxième *cette personne*, repris dans P1 par le pronom personnel *elle.* Si le texte contenait un prédicat transformateur, on pourrait *a priori* parler de référent évolutif et comprendre comment on passe d'une forme linguistique avec certaines caractéristiques de genre et de nombre à une autre forme linguistique, aux caractéristiques différentes. Mais lorsque le texte ne contient pas de prédicat transformateur, l'assignation du référent peut poser problème.

On peut rencontrer également des problèmes d'accessibilité du référent lorsque le choix des formes linguistiques des mentions est inapproprié. Par exemple, l'emploi d'une forme de haute accessibilité telle que le pronom suppose que le référent est connu de l'interlocuteur et saillant. Lorsque le référent n'a pas été convenablement introduit à l'aide de formes de moins haute accessibilité, le lecteur échoue à construire la référence.

Enfin, on constate des cas de discordance sémantique, le plus souvent avec le SN démonstratif *ce grand bruit*. Le référent introduit ne correspond alors pas au référent de la phrase imposée mais n'est pas suffisamment éloigné pour faire échouer complètement la référence. C'est le cas par exemple dans *un chat qui miaulait* : le lecteur comprend que l'élève a voulu prendre en compte la notion de bruit, mais l'expression choisie ne suffit pas pour produire un *grand* bruit.

Limites, apports et perspectives de recherche

Ainsi menée, notre recherche s'est limitée à l'étude d'un seul aspect de la cohésion textuelle, c'est-à-dire les marques de la continuité référentielles. La tâche-problème proposée, en imposant l'ordre et la nature des occurrences à traiter, s'éloigne des productions spontanées des élèves. On peut se demander jusqu'à quel point les résultats obtenus seraient confirmés par l'analyse de textes produits en réponse à une autre tâche d'écriture. Dans cette thèse, nous ne prenons pas en compte les caractéristiques de l'environnement didactique de chacune des classes, comme par exemple l'effet sur les stratégies référentielles des élèves des habitudes d'écriture plus ou moins développées selon les classes. D'autre part, mise à part la phrase fermoir qui est particulièrement difficile à traiter pour les élèves, nous avons vu que

l'occurrence *ce grand bruit* était majoritairement introduite à l'aide d'une phrase entière qui explique la provenance du bruit. On peut s'interroger sur les résultats obtenus si l'on modifiait l'ordre des constituants dans la phrase, par exemple *En entendant ce grand bruit, il se retourna.* La phrase pouvant être considérée comme introduisant une séquence descriptive, la phrase explicative aurait toute sa place après. Il serait intéressant de voir si les élèves introduisent tout de même le référent avant la phrase et par quelle forme linguistique.

D'un point de vue méthodologique, la posture du linguiste-didacticien qui aborde les textes d'élèves est particulière et peut constituer en soi une limite. Elle se situe entre le lecteur ordinaire qui est plus coopératif lorsqu'il s'agit de construire ou reconstruire une certaine cohérence, et le lecteur de textes d'élèves qui se positionne dans une attitude normative vis-à-vis du texte. À ce titre, le linguiste-didacticien peut s'interroger sur la part de calcul inférentiel qu'il est prêt à effectuer pour résoudre les problèmes de cohérence textuelle : doit-il considérer uniquement les textes dans lesquels la résolution des problèmes d'anaphore est nette, sans ambiguïté référentielle ? Peut-il prendre en compte les textes dans lesquels le lecteur doit effectuer des inférences, et jusqu'à quel point ? La réponse à ces questions peut influer sur le relevé des indices de résolution des problèmes référentiels.

Malgré ces limites, notre étude menée à son terme a permis de recueillir des données dont l'analyse alimente les travaux menés dans différents champs de recherche. En étudiant comment les élèves de différents niveaux scolaires introduisent et maintiennent les référents dans un texte, notre recherche apporte :

- Sur les plans linguistique et psycholinguistique : une cartographie des marques de compétences de la référence privilégiées par les élèves, en fonction de leur âge et du niveau de classe. En effet, en questionnant les procédés de continuité référentielle employés par les élèves dans des textes narratifs, dans une perspective de progressivité, nous avons mis au jour leurs compétences en matière de cohésion textuelle, en fonction de leur niveau de classe. En outre, le matériau collecté pour cette recherche constitue à la fois le corpus d'étude sur lequel s'appuie notre analyse, mais également, un enjeu secondaire de notre thèse puisqu'il s'inscrit dans un projet de constitution d'un « grand » corpus de texte scolaire en français langue première. Ce matériau est indispensable pour mettre au jour des invariants et des éléments de progressivité dans la mise en place des indicateurs de la cohésion textuelle. Il répond donc à un enjeu sociétal (Elalouf, 2011) car il peut

constituer à la fois une ressource pour les chercheurs mais aussi pour les formateurs d'enseignants et les étudiants se préparant aux métiers de l'enseignement.

- Sur le plan didactique : une meilleure compréhension du schéma de maîtrise de la référence par les élèves afin d'une part, de servir d'appui au jugement de cohérence des enseignants et à l'évaluation, et d'autre part, d'identifier des points d'entrée pour l'intervention didactique.

Sur les plans linguistiques et psycholinguistiques, le travail entrepris gagnerait à être approfondi. En effet, le travail mené dans le cadre cette recherche est principalement descriptif. En mettant en lumière les procédés de continuité référentielle employés par les élèves en fonction du niveau de classe, il fournit des éléments de compréhension du développement des compétences des élèves en matière de cohésion textuelle. L'analyse proposée n'est toutefois pas exhaustive. Il serait important de mettre en évidence les indices encore non exploités qui peuvent être le signe de différents degrés de maîtrise de la continuité référentielle. Par exemple, l'étude de la distance séparant les expressions référentielles d'une même occurrence, en fonction du niveau de classe, donnerait à voir la *texture* (Halliday & Hasan, 1976) plus ou moins importante des CDR, ce qui constituerait déjà en soi un indice supplémentaire de compétence des élèves : les élèves les plus âgés ont-ils des CDR plus denses et riches ? La distance entre les mentions pourrait également indiquer si les élèves construisent un véritable univers autour du référent avant d'insérer les phrases imposées, comme le feraient les scripteurs experts, ou bien s'ils juxtaposent seulement quelques phrases. On pourrait ensuite comparer les *textures* des différentes CDR appartenant à différentes entités de la consigne. Les CDR pourraient également faire l'objet d'une étude systématique concernant les relations qu'elles entretiennent entre elles (succession, entrecroisement, etc.), afin de voir si le niveau de classe des élèves influe sur le type de relation entre CDR. Un autre prolongement possible serait l'analyse du corpus via un outil d'annotation spécialisé tel qu'*ANALEC* par exemple. L'utilisation de ce type de logiciel, conçu pour des textes de scripteurs experts, permettrait peut-être de mettre en lumière des éléments nouveaux, si toutefois les spécificités des textes d'élèves permettent son utilisation. En effet nous l'avons vu, même transcrits, certains textes posent toujours des problèmes d'interprétation. En outre, on pourrait analyser et comparer les textes produits dans un même niveau de classe, par exemple en examinant le texte le plus faible et le texte le meilleur, ou encore comparer les

textes des élèves à des textes produits par des scripteurs experts, textes qui répondraient exactement à la même consigne. On envisagerait alors une progressivité plus grande encore, allant de l'enfant à l'adulte.

L'étude des ratures et des opérations d'écriture (ajout, suppression, remplacement, déplacement) pourrait apporter des indices quant aux modes de résolution des anaphores. En effet, en tant que traces d'élaboration des textes, elles sont un moyen de comprendre à la fois le cheminement du scripteur dans la prise en compte des problèmes de continuité référentielle, de cohésion et de cohérence textuelles, mais également, sa prise en compte de la compréhension du lecteur, montrant ainsi la capacité du scripteur à se décentrer.

D'un point de vue cognitif, il serait intéressant de procéder à un nouveau recueil de données avec la même consigne, mais en ajoutant un entretien orienté individuel avec les élèves après la rédaction, afin de permettre un retour sur leur expérience. L'entretien pourrait s'appuyer sur le texte produit par l'élève et chercherait à identifier les défauts de raisonnement qui ont conduit à des points de résistance tels que les ambiguïté référentielles par exemple, ou bien questionnerait les élèves sur les éléments qui leur ont posé le plus de difficultés et les comparerait en fonction du niveau de classe.

Des perspectives didactiques peuvent également être envisagées. En effet, la mise en exergue d'indicateurs de compétences des élèves en matière de cohésion textuelle et l'identification des problèmes de continuité référentielle, qui conduisent à un manque de cohérence globale du texte de l'élève, pourrait servir d'appui à deux éléments. D'une part, en venant compléter les quelques propositions didactiques issues de la recherche, elle pourrait permettre l'élaboration de séances d'apprentissages centrées sur l'acquisition de la cohésion textuelle, et plus particulièrement des marques de la continuité référentielle. Par exemple, en donnant à travailler aux élèves des textes dans lesquels il faut résoudre des anaphores concernant des entités exclusivement non humaines et non animées. Une séquence d'apprentissage pourrait également prendre appui sur la même consigne que celle de notre étude, en adaptant le nombre de phrases à insérer et leur complexité (nature des référents, présence d'anaphore résomptive…) en fonction du niveau de classe. Les textes produits pourraient faire l'objet de réécriture après des remarques de pairs ou un travail de sensibilisation au fonctionnement des CDR par exemple.

D'autre part, en pointant les éléments qui résistent, c'est-à-dire par exemple les problèmes d'interprétation référentielle que nous avons soulevés, la mise en exergue

d'indicateurs de compétences pourrait compléter les outils didactiques d'évaluation des compétences de scripteur des élèves. Cette perspective nous paraît intéressante dans la mesure où elle permettrait aux enseignants de dépasser les constats d'écart à la norme et d'englober à la fois les aspects locaux et globaux de l'organisation textuelle. L'élaboration d'un outil d'analyse donnerait à voir les compétences et connaissances en construction dans les productions d'élèves. Ce faisant, l'identification des besoins des élèves et des compétences prioritaires à travailler en fonction du niveau de classe en serait grandement facilitée.

Liste des Figures

Figure 1 : Schéma de la Givenness Hierachy, échelle de familiarité (Prince, 1981, repris et traduit en français par Gundel & al., 2000) 34
Figure 2 : Grille d'évaluation des écrits CLID, tirée de MAS, M. (1989) 144
Figure 3 : Grille d'évaluation avec indication de marques inadéquates et adéquates de cohérence, tirée de Carbonneau & Préfontaine (2005) 147
Figure 4 : extrait d'un texte d'élève avec l'annotation des écarts à la cohérence, tiré de Carbonneau et Préfontaine (2005) 147
Figure 5 : Grille d'évaluation avec indication de marques inadéquates et adéquates de cohérence, tirée de Carbonneau et Préfontaine (2005) 149
Figure 6 : Texte de Nina, CE2 L 191
Figure 7 : Texte de Nina, CE2 L 192
Figure 8 : Texte de Samuel, CM2 B 193
Figure 9 : Texte de Martin, CM1 T2 194
Figure 10 : Texte de Loulie, CE2 P 195
Figure 11 : Texte de Calixte, CM1 T1 198
Figure 12 : Texte d'Agathe, CE2 P 199
Figure 13 : Texte de Florentin, CE2 T 200
Figure 14 : Texte de Karelle, CE2 T 200
Figure 15 : Texte d'Aurore, CE2 P 201
Figure 16 : Texte de Romain, CE2 P 201
Figure 17 : Texte de Juliette, CE2 T 202
Figure 18 : Texte de Jules, CM2 B 203
Figure 19 : Texte de Laly, CE2 L 204
Figure 20 : Texte de Maëva, CM2 B 206
Figure 21 : Texte de Sérine, CM1 T2 207
Figure 22 : Texte de Karl, CE2 T 208
Figure 23 : Texte de Rayan, CM1 T2 209
Figure 24 : Transcription normée, Texte de Rayan, CM1 T2 210
Figure 25 : Texte de Romane, CE2 L 222
Figure 26 : Texte d'Estelle, CM1 T2 223
Figure 27 : Texte de Soliana, CM1 T1 225

Figure 28 : Texte de Romain, CE2 P 226
Figure 29 : Texte de Chemsa, CE2 U 228
Figure 30 : Texte de Suzanne, CM1 B 229
Figure 31 : Texte de Romane, CE2 L 231
Figure 32 : Texte d'Inès, CE2 U 237
Figure 33 : Texte de Karl, CE2 T 239
Figure 34 : Texte de Fethallah, CM1 B 239
Figure 35 : Texte de Thomas, CM1 T1 240
Figure 36 : Texte de Max, CM2 S 242
Figure 37 : Texte d'Adam, CE2 U 242
Figure 38 : Texte de Yassine, CM2B 244
Figure 39 : Texte de Loulie, CE2 P 245
Figure 40 : Texte de Victor, CE2 T 245
Figure 41 : Texte de Maxime, CM2 S 246
Figure 42 : Texte de Naélia, CM1 T1 246
Figure 43 : Texte d'Emma, CM1 T1 247
Figure 44 : Texte de Julian, CE2 L 248
Figure 45 : Texte de Antonin, CM2 S 249
Figure 46 : Texte de Lucas 2, CM1 T1 250
Figure 47 : Texte de Benjamin, CM2 T 250
Figure 48 : Texte de Quentin, CM2 M 252
Figure 49 : Texte de Nicolas, CE2 T 256
Figure 50 : Texte d'Arthur, CM1 T2 256
Figure 51 : Texte d'Emmaëlle, CE2 P 257
Figure 52 : Texte de Nicolas, CM2 T 259
Figure 53 : Texte de Matthieu, CM1 T1 262
Figure 54 : Texte de Rayene, CM1 B 267
Figure 55 : Texte de Mathias, CM2 M 269
Figure 56 : Texte de Thomas, CM1 T2 270
Figure 57 : Texte de Marine, CE2 T 271
Figure 58 : Texte d'Antoine, CM2 B 272
Figure 59 : Texte de Corentin, CM1 T1 273
Figure 60 : Texte de Léonore, CE2 T 273
Figure 61 : Texte de Neal, CM2 S 274

Figure 62 : Texte d'Alix, CM1 T1 275
Figure 63 : Texte de Mathéo, CM1 T1 277
Figure 64 : Texte de Naïla, CE2 P 278
Figure 65 : Texte d'Aliénor, CE2 C 278
Figure 66 : Texte de Bastien, CM2 B 278
Figure 67 : Texte de Soliana, CM1 T1 279
Figure 68 : Texte d'Olivia, CE2 U 281
Figure 69 : Texte de Loulie, CE2 P 281
Figure 70 : Texte de Noé, CM2 S 285
Figure 71 : Texte d'Alban, CE2 P 286
Figure 72 : Texte de Louise, CE2 P 288
Figure 73 : Texte d'Alix, CM1 T1 292
Figure 74 : Texte de Julian, CE2 L 297
Figure 75 : Texte d'Albane, CE2 T 299
Figure 76 : Texte de Camille, CM1 T1 299
Figure 77 : Texte de Nicolas, CM2 T 300
Figure 78 : Texte de Pauline, CM1 T1 300
Figure 79 : Texte d'Antoine, CM2 T 301
Figure 80 : Texte de Noémie, CE2 P 303
Figure 81 : Texte de Carlo, CM2 T 305
Figure 82 : Texte d'Olivia, CE2 U 306
Figure 83 : Texte de Soliana, CM1 T1 307
Figure 84 : Texte de Naélia, CM1 T 307
Figure 85 : Texte de Grégoire, CM2 B 309
Figure 86 : Texte de Jonas, CE2 T 310
Figure 87 : Texte de Carl, CM1 T1 314
Figure 88 : Texte d'Alban, CE2 P 317
Figure 89 : Texte de Mathieu, CM2 S 319
Figure 90 : Texte de Gabriel, CE2 U 325
Figure 91 : Texte de Max, CM2 S 327
Figure 92 : Texte d'Ingryd, CM2 B 327
Figure 93 : Texte d'Anass, CM1 B 328
Figure 94 : Texte de Naïla, CE2 P 328
Figure 95 : Texte d'Olivia, CE2 U 329

Figure 96 : Texte de Mathéo, CM1 T1 330
Figure 97 : Texte de Lucile, CM1 T1 330
Figure 98 : Texte de Mathis, CM1 T1 331
Figure 99 : Texte de Juliette, CE2 T 333
Figure 100 : Texte de Nathan, CM2 S 333
Figure 101 : Texte d'Eliott 2, CE2 T 335
Figure 102 : Texte de Rose, CE2 U 339
Figure 103 : Texte de Mathéo, CM1 T1 340
Figure 104 : Texte de Karl, CM1 T1 342
Figure 105 : Texte de Max, CM2 S 343
Figure 106 : Texte de Jules, CM2 B 346
Figure 107 : texte de Cindy, CM1 T1 353
Figure 108 : Texte de Jolin, CE2 T 354
Figure 109 : Texte de Lisa, CM2 T 355
Figure 110 : Texte de Nina, CE2 L 357
Figure 111 : Texte sans nom 1, CM2 S 358
Figure 112 : Texte d'Inaya, CM2 B 360
Figure 113 : Texte d'Amélie, CE2 P 361
Figure 114 : Texte de Lana, CM2 M 363
Figure 115 : Texte d'Elouan, CM1 T1 371
Figure 116 : Texte d'Enzo, CM2 S 375
Figure 117 : Texte d'Ange, CE2 P 376
Figure 118 : Texte de Juliette, CE2 T 377
Figure 119 : Texte de Solène, CM1 T2 377
Figure 120 : Texte d'Enzo, CE2 P 378
Figure 121 : Texte de Salomé, CM2 B 379
Figure 122 : Texte de Quentin, CM2 M 380
Figure 123 : Texte d'Enzo, CE2 P 380
Figure 124 : Texte de Noëmie, CM2 B 383
Figure 125 : Texte d'Ugo, CM2 M 385
Figure 126 : Texte d'Enzo, CE2 P 386
Figure 127 : Texte de Naïla, CE2 P 388

Liste des Tableaux

Tableau 1 : Nombre de textes recueillis par classes et niveau scolaire 188
Tableau 2 : Synthèse des conventions de codage utilisées pour la transcription diplomatique du corpus 196
Tableau 3 : Localisation de la source du pronom personnel elle 236
Tableau 4 : Formes de la première mention du référent du pronom personnel elle 238
Tableau 5 : Formes lexicales de la première mention du référent du pronom personnel elle 241
Tableau 6 : Nombre d'expansions dans les formes de la première mention du référent du pronom personnel elle 243
Tableau 7 : Nombre d'expansions dans les formes de la première mention du référent du pronom personnel elle 244
Tableau 8 : Type de deuxième expansion dans les formes de la première mention du référent du pronom personnel elle 247
Tableau 9 : Composition des propositions relatives présentes dans la deuxième expansion de la première mention du référent du pronom personnel elle 249
Tableau 10 : Types de fonctions syntaxiques occupées par la première mention du référent du pronom personnel elle 252
Tableau 11 : Nombre de mentions dans la CDR du pronom personnel elle 254
Tableau 12 : Formes de la deuxième mention du référent du pronom personnel elle dans les CDR minimales 258
Tableau 13 : Fonctions syntaxiques de la deuxième mention du référent du pronom personnel elle dans les CDR minimales 260
Tableau 14 : Localisation de la source du référent du pronom personnel il 266
Tableau 15 : Formes de la première mention du référent du pronom personnel il 268
Tableau 16 : Nombre d'expansions dans les formes de la première mention du référent du pronom personnel il 276
Tableau 17 : Type de première expansion dans les formes de la première mention du référent du pronom personnel *il* 276
Tableau 18 : Différents types de fonctions syntaxiques prises par le SN de la première mention du référent du pronom personnel *il* 280
Tableau 19 : Nombre de mentions dans la CDR incluant le pronom personnel il 283

Tableau 20 : Formes de la deuxième mention du référent du pronom personnel il dans les CDR minimales 287
Tableau 21 : Fonctions syntaxiques de la deuxième mention du référent du pronom personnel il dans les CDR minimales 289
Tableau 22 : Localisation de la source du SN cette maison 296
Tableau 23 : Formes de la première mention du référent du SN démonstratif cette maison 298
Tableau 24 : Nombre d'expansions de la première mention du référent du SN démonstratif cette maison 304
Tableau 25 : Type de première expansion dans les formes de la première mention du référent du SN démonstratif cette maison 304
Tableau 26 : Type de deuxième expansion dans les formes de la première mention du référent du SN démonstratif cette maison 306
Tableau 27 : Différents types de fonctions syntaxiques prises par la première mention du référent du SN démonstratif cette maison 309
Tableau 28 : Nombre de mentions dans la CDR incluant le référent du SN démonstratif cette maison 312
Tableau 29 : Formes de la première mention du référent du SN démonstratif cette maison dans les cas de cataphore (P1 + 1 mention) 314
Tableau 30 : Formes de la deuxième mention du référent du SN démonstratif cette maison dans les CDR minimales 315
Tableau 31 : Fonctions syntaxiques de la deuxième mention du référent du SN démonstratif cette maison dans les CDR minimales 316
Tableau 32 : Localisation de la source du SN démonstratif ce grand bruit 324
Tableau 33 : Formes de la première mention du référent du SN démonstratif ce grand bruit 326
Tableau 34 : Nombre d'expansions de la première mention du référent du SN démonstratif ce grand bruit 332
Tableau 35 : Type de première expansion dans les formes de la première mention du référent du SN démonstratif ce grand bruit 332
Tableau 36 : Différents types de fonctions syntaxiques prises par la première mention du référent du SN démonstratif ce grand bruit 334
Tableau 37 : Nombre de mentions dans la CDR incluant le SN démonstratif ce grand bruit 336

Tableau 38 : Formes de la première mention du référent du SN démonstratif ce grand bruit dans les cas de cataphore (P2 + 1 mention après P2) 339
Tableau 39 : Formes de la deuxième mention du référent du SN démonstratif ce grand bruit dans les CDR minimales 341
Tableau 40 : Fonctions syntaxiques de la deuxième mention du référent du SN démonstratif ce grand bruit dans les CDR minimales 343
Tableau 41 : Localisation de la source du SN défini les enfants 351
Tableau 42 : Types de référents quand P3 est la seule mention du SN défini les enfants 356
Tableau 43 : Formes de la première mention du référent du SN défini les enfants 357
Tableau 44 : Différents types de fonctions syntaxiques prises par la première mention du référent du SN défini les enfants 363
Tableau 45 : Nombre de mentions dans la CDR incluant le référent du SN défini les enfants 364
Tableau 46 : Formes de la première mention du référent du SN défini les enfants dans les cas d'anaphore (1 mention avant P3 + P3) 367
Tableau 47 : Fonctions syntaxiques de la première mention du référent du SN défini les enfants dans les cas d'anaphore (1 mention avant P3 + P3) 367
Tableau 48 : Localisation de la source du SN défini la nuit 373
Tableau 49 : Formes de la première mention du référent du SN défini la nuit 374
Tableau 50 : Types de fonctions syntaxiques prises par la première mention du référent du SN défini la nuit 380
Tableau 51 : Étude de la longueur moyenne des CDR et nombre de mentions dans la CDR la plus longue, pour chaque occurrence 397
Tableau 52 : Taux d'échec de l'introduction des référents dans le discours 402
Tableau 53 : Nombre d'expressions référentielles intégrées dans le texte, en fonction du niveau de classe 406

INDEX

anaphore, **60**
anaphore associative, **72**
anaphore fidèle, **70**
anaphore hyperonymique, **71**
anaphore infidèle, **71**
anaphore pronominale, **73**
anaphore résomptive, **72**
anaphores conceptuelles, **71**
ancienne information, **31**
cataphore, **63**
chaîne de référence, **75**
cohésion lexicale, **43**
collocation, **44**
conjonction, **42**
coréférentialité, **62**
corpus, **157**
décontextualisation, **102**
dénominations passe-partout, **101**
distance inter-maillonnaire, **76**
ellipse, **41**
endophorique, **47**
évaluation, **141**
exophorique, **47**
genre discursif, **82**
genre narratif, **83**
givenness hierarchy, **32**
information nouvelle, **31**
jugement de cohérence, **150**
knowledge telling, **113**
marques de cohésion, **39**
mémoire discursive, **25**
méta-règle de non-contradiction, **28**
méta-règle de progression, **28**
méta-règle de relation, **29**
méta-règle de répétition, **27**
méta-règles de cohérence, **27**
mise en texte, **111**
planification, **110**
progression thématique, **35**
redénomination, **77**
référence, **45**
représentation mentale, **24**
révision, **111**
saillance, **36**
stratégie de non-changement de rôle, **92**
stratégie de proximité ou de distance minimale, **92**
stratégie du sujet thématique, **98**
stratégie lexicale, **92**
substitution, **40**
syllepse, **73**
tâche-problème, **184**
théorie de l'accessibilité, **34**
transcription chronologique, **171**
transcription diplomatique, **170**
transcription linéarisée, **170**

BIBLIOGRAPHIE

ADAM, J-M. (2005). *La linguistique textuelle. Introduction à l'analyse textuelle des discours.* Paris : A. Colin.

ADAM, J-M. (2012). Discursivité, généricité et textualité. Distinguer pour penser la complexité des faits de discours. *Recherches*, n°56, [En ligne].

AEBY, S., DE PIETRO, J-F. & WIRTHNER, M. (2001). *Français 2000. Dossier préparatoire. L'enseignement du français en Suisse romande : un état des lieux et des questions.* Neuchâtel : IRDP.

ALONSO RAMOS, M. (1994). Hacia una definición del concepto de colocación: de J.R. Firth a I.A. Mel'čuk. *Revista de Lexicografía*, I, 9-28.

APOTHÉLOZ, D. (1995a). *Rôle et fonctionnement de l'anaphore dans la dynamique textuelle.* Genève : Droz.

APOTHÉLOZ, D. (1995b). Nominalisations, référents clandestins et anaphores atypiques. *TRANEL*, 143-173.

APOTHELOZ, D. & PEKAREK DOEHLER, S. (2003). Nouvelles perspectives sur la référence : des approches informationnelles aux approches interactionnelles. *Verbum*, n°15, 109-136.

ARIEL, M. (1988). Referring and accessibility. *Journal of Linguistics,* n°24, 65-87.

ARIEL, M. (1990). *Accessing Noun-Phrase Antecedents*. London: Routledge.

ARIEL, M. (2001). Accessibility theory: an overview. *In* Sanders, T., Schilperoord, J. & Spooren, W. (eds.), *Text representation: linguistic and psycholinguistic aspects*. Amsterdam/Philadelphia: John Benjamins, 29-88.

AURIAC-SLUSARCZYK, E., CHANQOY, L., COGIS, D., GARCIA-DEBANC, C., GUNNARSSON, C., LARGY, P., LEBLAY, C., SLUSARCZYK, B. (2008). Étude pluridisciplinaire du processus de révision chez de jeunes rédacteurs du primaire au collège. *In* Alamargot, D., Bouchand, J., Lambert, E., Millogo, V. & Beaudet, C. (éds.), *Proceedings of the International Conference « De la France au Québec : l'Ecriture dans tous ces états »,* novembre 2008, Poitiers, France.

BALLY, C. (1909). *Traité de stylistique française.* Paris : Klincksieck.

BAMBERG, M. (1986). A functional approach to the acquisition of anaphoric relationships. *Linguistics*, n°24, 227-284.

BARTLETT, E.J. (1984). Anaphoric Reference in Written Narratives of Good and Poor Elementary School Writers. *Journal of Verbal Learning and Verbal Behavior*, n°23, 540-552.

BEAUDET, C. (2001). Clarté, lisibilité, intelligibilité des textes : un état de la question et une proposition pédagogique. *Recherches en rédaction professionnelle*, vol.1, 1-17.

BENVENISTE, E. (1966). *Problèmes de linguistique générale*. T1, Paris : Gallimard.

BENVENISTE, E. (1974). *Problèmes de linguistique générale*. T2, Paris : Gallimard.

BEREITER, C. & SCARDAMALIA, M. (1987). *The Psychology of Written Composition*. Hillsdale, NJ : LEA.

BERGERON, R. (2003). Relire son texte pour détecter et résoudre des problèmes de cohérence : un défi de taille pour les scripteurs en apprentissage. *Québec français*, n°128, 55-57.

BERGERON, R. & BLAIN, R. (2003). La cohérence textuelle. Présentation. *Québec français*, n°128, 47-48.

BERMAN, R. (2004). Between emergence and mastery: The long developmental route of language acquisition. *In Language Development across Childhood and Adolescence.* Trends in Language Acquisition Research, Vol 3, Amsterdam: John Benjamins, 9-34

BERMAN, R. (2008). The psycholinguistics of developing text construction. *Journal of Child Language*, n°35, 735-771.

BERMAN, R. (2014). Cross-linguistic comparisons in child language research. *Journal of Child Language*, n°41, 26-37.

BERMAN, R. & KATZENBERGER, I. (2004). Form and function in introducing narrative and expository texts: a developmental perspective. *Discourse Processes*, 38(1), 57-94.

BERRENDONNER, A. (1983). Connecteurs pragmatiques et anaphore. *Cahiers de linguistique française*, n°5, 215-246.

BERRENDONNER, A. (1986). *Référence et mémoire discursive*. Colloque du Centre de Recherches Sémiologiques de l'Université de Neuchâtel, octobre 1986, Neuchâtel, Suisse.

BERRENDONNER, A. & REICHLER-BEGUELIN, M-J. (1989). Décalages : les niveaux de l'analyse linguistique. *Langue Française*, n°81, 99-125.

BERRENDONNER, A., MOSEGAARD HANSEN, M-B, ZAFIU, R. (2016). Section 10 : Linguistique textuelle et analyse du discours. *Actes du XXVIIe Congrès international de linguistique et de philologie romane,* Juin 2013, Nancy, France.

BIBER, D. (1988). *Variation across speech and writing.* Cambridge: Cambridge University Press.

BILGER, M. & CAPPEAU, P. (2013). Comment les données de corpus pourraient renouveler les manuels de grammaire ? *Linx*, n°68-69, 177-199.

BLANCHE-BENVÉNISTE, C. (1997). *Approches de la langue parlée en français*. Paris : Ophrys.

BLANCHE-BENVÉNISTE, C. & CHERVEL, A. (1966). Recherches sur le syntagme substantif. *Cahiers de lexicologie*, n°9, 3-37.

BLASS, R. (1990). *Relevance Relations in Discourse. A study with special reference to Sissala*. Cambridge: Cambridge University Press.

BONNET, C. (1994). *Plume en main ou l'itinéraire de l'enfant qui apprend à écrire*. Lausanne : CVRP.

BONNET, C., CORBLIN, C., ELALOUF, M-L, BARHOUMI, J. (1998). *Les procédés d'écriture chez les élèves de 10 à 13 ans, un stade de développement*. Lausanne : Éditions LEP.

BORÉ, C. (2007). Corpus et genres scolaires : affinités, difficultés. *Le français aujourd'hui,* n°159, 19-28.

BORE, C. (2011). L'énonciation des brouillons et la question du sujet scolaire. *Pratiques,* 149-150, 71-90.

BORÉ, C. & ELALOUF, M-L. (2017). Deux étapes dans la construction de corpus scolaires : problèmes récurrents et perspectives. *Corpus*, n°16, 31-64.

BOSCH, P. (1983). *Agreement and Anaphora. A study of the role of Pronouns in Syntax and Discourse*. London: Academic Press.

BOURDIN, B. & FAYOL, M. (2000). Is graphic activity cognitively costly? A developmental approach. *Reading and writing*, n°13, 183-196.

BOURDIN, B. & FAYOL, M. (2002). Even in adults, written production is still more costly than oral production. *International journal of Psychology*, n°37, 219-227.

BOWKER, L. & PEARSON, J. (2002). *Working with specialized language, a practical guide to using corpora.* London: Routledge.

BRANCA-ROSOFF, S. (1999). Types, modes et genres : entre langue et discours. *Langage et société*, n°87, 5-40.

BRISSAUD, C. & GROSSMANN, F. (2009). La construction des savoirs grammaticaux. *Repères*, n°39, 5-15. [En ligne].

BRONCKART, J-P. (1996). *Activité langagière, textes et discours. Pour un interactionnisme socio-discursif.* Lausanne-Paris : Delachaux et Niestlé.

BRONCKART, J-P. (1997). *Activité langagière, textes et discours, Pour un interactionnisme socio-discursif.* Genève-Paris : Droz

BRONCKART, J-P. (2016). Que faire de la grammaire et comment en faire ? *Pratiques*, n°169-170, [En ligne].

BRONCKART, J-P., BAIN, D., SCHNEUWLY, B. & *al*. (1985). *Le fonctionnement des discours.* Neuchâtel et Paris : Delachaux et Niestlé.

BRONCKART, J-P. & SCHNEUWLY, B. (1991). Children's production of textual organizers. *In* Piéraut-Le Bonnec, G. & Dolitsky, M. (eds.), *Language bases…discourse bases*, Amsterdam: Benjamins.

BROWN, R. (1973). *A first Language.* London: George Allen & Unwin Ltd.

BROWN, G. & YULE, G. (1983). *Discourse analysis.* Cambridge: Cambridge University Press.

BUCHETON, D. (1994). Décloisonner ou comment mettre la langue, le langage, le texte à distance. *Le français aujourd'hui,* n°107, 60-65.

BULEA BRONCKART, E. & BRONCKART, J-P. (2017). La place de la grammaire dans l'enseignement du français. La situation romande et ses implications pour la formation des enseignants. *In* Bulea Bronckart et Gagnon (dir.), *Former à l'enseignement de la grammaire*. Presses universitaires du Septentrion.

CAPPEAU, P. (2013). Quelques propriétés des sujets lexicaux. Remarques et analyse des récits d'élèves (dans leur version initiale). In *Écriture et réécritures chez les élèves. Un seul corpus, divers genres discursifs et méthodologies d'analyse*. Sciences du langage : Carrefours et points de vue n°10, Louvain-La-Neuve : Academia L'Harmattan.

CAPPEAU, P. & ROUBAUD, M-N. (2005). *Enseigner les outils de la langue avec les productions d'élèves.* Paris : Bordas.

CARBONNEAU, C. & PREFONTAINE, C. (2005). Enseigner et évaluer la cohérence textuelle, *Québec français*, n°138, été 2005, 78-81.

CARON, J. (1989). *Précis de psycholinguistique*. Paris : PUF.

CAVANAGH, M. (2006). Un programme pour soutenir les élèves dans l'apprentissage de la cohérence textuelle. *Québec français,* n°141, 94-96.

CHABANNE, J-C. & BUCHETON, B. (2002). *Ecrire en ZEP*. Paris : Delagrave.

CHAFE, W. (1987). Cognitive constraints on information flow. *In* Tomlin, R. (ed.), *Coherence and grounding in discourse*, Amsterdam/Philadelphia: John Benjamins, 24-51.

CHAFE, W. (1994). *Discourse, consciousness, and time: the flow and displacement of conscious experience in speaking and writing.* Chicago : University of Chicago Press.

CHANQUOY, L. & FAYOL, M. (1991). Étude de l'utilisation des signes de ponctuation et des connecteurs chez des enfants (8-10 ans) et des adultes. *Pratiques*, n°70, 107-124.

CHAROLLES, M. (1978). Introduction aux problèmes de la cohérence des textes. *Langue française,* n°38, 7-41.

CHAROLLES, M. (1988a). Les plans d'organisation textuelle périodes, chaînes, portées et séquence. *Pratiques,* n°57, 3-13

CHAROLLES, M. (1988b). Les études sur la cohérence et la connexité textuelles depuis la fin des années 1960. *Modèles Linguistiques*, tome X, fasc. 2, n°20, 45-66.

CHAROLLES, M. (1988c). La gestion des risques de confusion entre personnage dans une tâche rédactionnelle. *Pratiques*, n°60, 75-97.

CHAROLLES, M. (1990). L'anaphore associative. Problèmes de délimitation. *Verbum*, n°13-3, 119-148.

CHAROLLES, M. (1995). Cohésion, cohérence et pertinence du discours. *Travaux de Linguistique,* n°29, 125-151.

CHAROLLES, M. (1997). L'encadrement du discours : univers, champs, domaines et espaces. *Cahier de Recherche Linguistique*, n°6, 1-73.

CHAROLLES, M. (2001). Référents évolutifs et évolution de la référence. *In* De Mulder, W. & Schnedecker, C., *Les référents évolutifs entre linguistique et philosophie,* Paris : Klincksieck, 39-97.

CHAROLLES, M. (2002). *La référence et les expressions référentielles en français*. Paris : Ophrys.

CHAROLLES, M. (2011). Cohérence et cohésion du discours. *In* Holker, K. ; Marello, C. *Dimensionen der Analyse Texten und Diskursiven – Dimensioni dell'analisi di testi e discoursi, Lit Verlag*, 153-173.

CHAROLLES, M. (2014). Annotation des expressions référentielles et profondeur de traitement. In M. Fossard & M-J Béguelin (éds.). *Nouvelles perspectives sur l'anaphore.* Peter Lang : Berne, 55-98.

CHAROLLES, M. & STORME, B. (2015). La référence aux personnages et aux lieux dans *L'Occupation des sols*. *Revue Sciences/Lettres*, n°3, 1-21.

CHARTRAND, S. (2008). *Progression dans l'enseignement du français langue première au secondaire québécois.* Québec : Les publications Québec français.

CHARTRAND, S. (2011). Prescriptions pour l'enseignement de la grammaire au Québec. Quels effets sur les pratiques ? *Le français aujourd'hui,* n°173, 45-54.

CHARTRAND, S. & PARET, M-C. (2010). Rénover l'enseignement grammatical au Québec. Les tenants et les aboutissants du programme de 1995. *Québec français*, n°156, 62-65.

CHASTAIN, C. (1975). Reference and Context. *In Language Mind and Knowledge*, Minneapolis: University of Minnesota Press, 194-269.

CHERVEL, A. (1988). L'histoire des disciplines scolaires : réflexions sur un domaine de recherche. *Histoire de l'Éducation*, n°38, 59-119.

CHERVEL, A. (2006). *Histoire de l'enseignement du français du XVIIe au XIXe siècle*. Paris : Retz.

CHESNET, D. & ALAMARGOT, D. (2005). Analyse en temps réel des activités oculaires et graphomotrices du scripteur : intérêts du dispositif « eye and pen ». *L'année psychologique*, n°105, 477-520.

CHINCHOR, N. & HIRSCHMANN, L. (1997). MUC-7 Conference Task Definition, Version 3.0, *Proceedings of MUC-7.*

CHUY, M. & RONDELLI, F. (2010). Traitement des contraintes linguistiques et cognitives dans la construction de la cohérence textuelle. *Langages*, n°177, 83-111.

CLARK, H. & HAVILAND, S. (1977). Comprehension and the Given-New contract. *In* Freedle, R. (eds.), *Discourse production and comprehension*, 1-40, Norwood, NJ: Ablex.

COMBETTES, B. (1977). Ordre des éléments de la phrase et linguistique du texte. *Pratiques*, n°13, 91-101.

COMBETTES, B. (1983). Pour une grammaire textuelle. Bruxelles : De Boeck-Duculot.

COMBETTES, B. (1986). Introduction et reprise des éléments d'un texte. *Pratiques,* n°49, 69-84.

COMBETTES, B. (1992). Hiérarchie des référents et connaissance partagée : les degrés dans l'opposition connu/nouveau. *L'information grammaticale,* n°54, 11-14.

COMBETTES, B. (1995). Système de la langue et linguistique du texte. *Le français dans tous ses états,* n°30.

COMBETTES, B. (2016). La « grammaire de texte » dans les textes officiels depuis le Plan de rénovation. *Pratiques*, n°169-170, [En ligne].

COMBETTES, B. & TOMASSONE, R. (1988). Le texte informatif, aspects linguistiques. Bruxelles : De Boeck Université.

COMBETTES, B. & CHAROLLES, M. (1999). Contribution pour une histoire récente de l'analyse du discours. *Langue française*, n°121, 76-116.

CONDAMINES, A. (2005). Anaphore nominale infidèle et hyperonymie : le rôle du genre textuel. *Revue de Sémantique et Pragmatique*, n°18, 33-52.

CORBLIN, F. (1985). Remarques sur la notion d'anaphore. *Revue Québécoise de linguistique, vol. 15,* n°1, 173-195.

CORBLIN, F. (1987a). Les chaînes de référence naturelles. *T.A. Information,* vol. 1, 5-21.

CORBLIN, F. (1987c). *Indéfini, défini et démonstratif. Constructions linguistiques de la référence.* Genève : Droz.

CORBLIN, F. (1995). *Les formes de reprise dans le discours. Anaphores et chaînes de référence*. Presses Universitaires de Rennes.

CORNISH, F. (1987). Anaphoric pronouns: under linguistic control or signaling particular discourse representation? *Journal of Semantics*, n°5, 233-260.

CORNISH, F. (1990). Anaphore pragmatique, référence et modèle du discours. *In* Kleiber, G. et Tyvaert, J-E (éds.), *L'anaphore et ses domaines*, Paris : Klincksieck, 123-138.

CORNISH, F. (1999). *Anaphora, discourse and understanding. Evidence from English and French*. Clarendon: Oxford Linguistics.

CORNISH, F. (2001). "Modal" that as determiner and pronoun: the primacy of the cognitive-interactive dimension. *English Journal and Linguistics, 5(2),* 297-315.

CORNISH, F. (2009). Le rôle des anaphores dans la mise en place des relations de cohérence dans le discours : l'hypothèse de J.R Hobbs. *Journal of French Language Studies,* n°19 (2), 159-181.

CORNISH, F. (2014). Indexicals and context: context-bound prerequisite(s), ongoing processing and aftermaths of the discourse referring act. *In* Fossard, M. & Reichler-Beguelin, M.J. (eds.), *Nouvelles perspectives sur l'anaphore*, Bern : Peter Lang.

DALBERA, J-P. (2002). Le corpus entre données, analyse et théorie. *Corpus,* n°1, [En ligne].

DAVID, J. (2000a). Étudier les textes d'enfants : revue de travaux. *In* Fabre-Cols, C. (dir.), *Apprendre à lire des textes d'enfants.* Bruxelles : Boeck-Duculot, 274-288.

DAVID, J. (2000b). Le lexique et son acquisition : aspects cognitifs et linguistiques. *Le français aujourd'hui*, n°131, 31-40.

DAVID, J. & DOQUET, C. (2016). Les écrits d'élèves : un corpus de référence pour le français contemporain. *Congrès Mondial de Linguistique Française, CMLF 16, juillet 2016*, Tours, France.

DEDEYAN, A., LARGY, P. & NEGRO, I. Mémoire de travail et détection d'erreurs d'accord verbal : étude chez le novice et l'expert. *Langages*, n°164, 57-70.

DE MULDER, W. (1997). Les démonstratifs : des indices de changement de contexte. *In* Flaux, N., Van de Velde, D. & De Mulder, W., (eds), 137-193.

DE MULDER, W. (2000). Démonstratifs et accessibilité. *Verbum, XXII, 1,* 103-125.

DEUTSCH, W., KOSTER, C., KOSTER, J. (1986). What can we learn from children's errors in understanding anaphora? *Linguistics*, n°24, 203-225.

DE VOGUE, S., ESPINOZA, N., GARCIA, B., PERINI, M., SITRI, F. & WATOREK, M. (2017). Constitution d'un grand corpus émergents et novices : principes et méthodes. *Corpus*, n°16, 65-85.

DE WECK, G. (1991). *La cohésion dans des textes d'enfants. Étude du développement des processus anaphoriques.* Neuchâtel et Paris : Delachaux et Niestlé.

DE WECK, G. & SCHNEUWLY, B. (1994). Anaphoric Procedures in Four Text Types Written by Children. *Discourse Processes*, n°17, 101-113.

DJEBBOUR, S. & LARTIGUE, R. (1994). Anaphores et désignation des personnages dans le récit au cycle 3. *Repères*, n°9, 119-132.

DOQUET, C. (2013). Dépasser l'impossible alliance : quelles interactions entre production écrite et maîtrise de la langue ? *Le français aujourd'hui*, n°2, 119-130.

DOQUET, C., DAVID, J. et FLEURY, S. (2017a). Spécificités et contraintes des grands corpus de textes scolaires : problèmes de transcription, d'annotation et de traitement. Introduction. *Corpus*, n°16, 7-13.

DOQUET, C., ENOIU, V., FLEURY, S. & MAZIOTTI, S. (2017b). Problèmes posés par la transcription et l'annotation d'écrits d'élèves. *Corpus,* n°16, 133-155.

ELALOUF, M-L. (2004). Constitution d'un grand corpus de textes d'élèves. Problèmes méthodologiques et premiers résultats. *Linx*, n°51, 129-146.

ELALOUF, M-L. (2011). Constitution de corpus scolaires et universitaires : vers un changement d'échelle ? *Pratiques,* n°149-150, 56-70.

ELALOUF, M-L. (2016). L'analyse linguistique des textes d'élèves au travers des annotations, 1982-2014. *Pratiques*, n°169-170, [En ligne].

ELALOUF, M-L. (dir.), BERTUCCI, M-M., BORE, C. & *al.* (2005). *Ecrire entre 10 et 14 ans : un corpus, des analyses, des repères pour la formation.* Versailles : CRDP de l'académie de Versailles.

ELALOUF, M-L. & BORÉ, C. (2007). Construction et exploitation de corpus d'écrits scolaires. *Revue française de la linguistique appliquée*, vol. XII, 53-70.

FABRE-COLS, C. (1990). *Les brouillons d'écoliers ou l'entrée dans l'écriture.* Grenoble : Ceditel & L'atelier du texte.

FABRE-COLS, C. (2000). *Apprendre à lire des textes d'enfants.* Bruxelles : De Boeck-Duculot.

FABRE-COLS, C. (2002). *Réécrire à l'école et au collège, de l'analyse des brouillons à l'écriture accompagnée*. Paris : ESF.

FALL, K., VIGNAUX, G. & BOIVIN, M. (1994). Les études sur l'anaphore : quelques orientations actuelles dans le contexte européen. *Linguistica atlantica,* n°16, 45-58.

FAVART, M. (2005). Les marques de cohésion : leur rôle fonctionnel dans l'acquisition de la production écrite de texte. *Psychologie française,* n°50, 305-322. .

FAVART, M. & OLIVE, T. (2005). Modèles et méthodes d'étude de la production écrite. *Psychologie française*, n°50, 273-285.

FAVART, M. & CHANQUOY, L. (2007). Les marques de cohésion comme outils privilégiés de la textualisation : une comparaison entre élèves de CM2 et adultes experts. *Langue française*, n°155, 51-68.

FAYOL, M. (1984). L'approche cognitive de la rédaction : une perspective nouvelle. *In Repères pour la rénovation de l'enseignement du français*, n°63, 65-69.

FAYOL, M. (2002). *Production du Langage*. Paris : Lavoisier.

FAYOL, M. & SCHNEUWLY, B. (1987). La mise en texte et ses problèmes. *In* Chiss, J-L., Laurent, J-P., Meyer, J-C., Romian, H. & Schneuwly, B. *Apprendre/enseigner à produire des textes écrits.* Actes du 3e colloque international de didactique du français, septembre 1986, Namur. Bruxelles : De Boeck, 223-239.

FIRTH, J.R. (1957). *Papers in linguistics, 1934-1951*. London: Oxford University Press.

FRASER, T. & JOLY, A. (1980). Le système de la deixis. Endophore et cohésion discursive en anglais. In *Modèles linguistiques*, II, 2, 22-51.

FORT, K. (2012). *Les ressources annotées, un enjeu pour l'analyse de contenu : vers une méthodologie de l'annotation manuelle de corpus*. Thèse de doctorat de l'université de Paris 13, soutenue le 7 décembre 2012.

FOSSARD, M. (2008). Aspects linguistiques et cognitifs en jeu lors du traitement d'expressions démonstratives anaphoriques. Congrès Mondial de Linguistique Française, CMLF 08, juillet 2008, Paris, France.

GALMICHE, M. (1986). Notes sur les noms de masse et le partitif. *Langue Française*, n°72, 40-53.

GAONAC'H, D. & LARIGAUDERIE, P. (2000). *Mémoire et fonctionnement cognitif : la mémoire de travail*. Paris : Armand Colin.

GARCIA-DEBANC, C. (1990). *L'élève et la production d'écrits*. Metz : CASUM.

GARCIA-DEBANC, C. (1993). Enseignement de la langue et production d'écrits. *Pratiques*, n°77, 3-23.

GARCIA-DEBANC, C. (2010). Segmentation, connexion et indexation dans des productions écrites d'élèves de 9 à 13 ans de deux genres textuels. *Synergies Pays Scandinaves*, n°5, 81-96.

GARCIA-DEBANC, C. & BONNEMAISON, K. (2014). La gestion de la cohésion textuelle par des élèves de 11-12 ans : réussites et difficultés. *Congrès Mondial de Linguistique Française, CMLF 14, juillet 2014*, Berlin, Allemagne.

GARCIA-DEBANC, C. & BRAS, M. (2016). Vers une cartographie des compétences de cohérence et de cohésion textuelle dans une tâche-problème de production écrite réalisée par des élèves d e9 à 12 ans : indicateurs de maîtrise et progressivité. *In* Plane, S., Bazerman, C., Rondelli, F., Donahue, C. et *al.* (éd.), *Recherches en écriture : regards pluriels,* Actes du colloque Writing Research Across Boarders 2014, *Recherches Textuelles* n°13, Metz : Université de Lorraine, 39-62.

GARCIA-DEBANC, C., HO-DAC, L-M., BRAS, M. & REBEYROLLE, J. (2017). Vers l'annotation discursive de textes d'élèves. *Corpus,* n°16, 157-184.

GARDENT, C. & MANUÉLIAN, H. (2005). Création d'un corpus annoté pour le traitement des descriptions définies. *L'objet*, vol. 8, n°2, 1-15.

GARROD, S. & SANFORD, A.J. (1982). The Mental Representation of Discourse in a Focussed Memory System: Implication for the Interpretation of Anaphoric Noun Phrases, *Journal of Semantics*, 1, 21-41.

GARROD, S. & SANFORD, A.J. (1988). Thematic subjecthood and cognitive constraints on discourse structure. *Journal of Pragmatics*, n°12, 519-534.

GEOFFRE, T. (2013). *Vers le contrôle orthographique au cycle 3 de l'école primaire*, thèse de doctorat de Sciences du Langage, spécialité Didactique et Linguistique, Université de Grenoble.

GERNSBACHER, M. & GIVON, T. (1995). Introduction, coherence as a mental entity. *In* Gernsbacher, M., Givón, T. (eds.), *Coherence in spontaneous text*, n°VII-X, Philadelphia: John Benjamins.

GIVON, T. (1983). Topic continuity in discourse: An introduction. *In* T. Givón (ed.), *Topic Continuity in Discourse. A quantitative cross-language study*, Amsterdam/Philadelphia: John Benjamins Publishing, 1-41.

GIVON, T. (1989). *Mind, Code, and Context: Essays in Pragmatics*. Hillsdale (NJ): Lawrence Erlbaum Associate.

GIVON, T. (1992). The grammar of referential coherence as mental processing instructions. *Linguistics*, 30/1, 5-55.

GIVON, T. (1995a). Coherence in text *vs.* Coherence in mind. *In* Gernsbacher, M., Givón, T. (eds.), *Coherence in spontaneous text*, Philadelphia: John Benjamins, 59-116.

GIVON, T. (1995b). *Functionalism and grammar*. Amsterdam/Philadelphia: Jhon Benjamins.

GRAESSER, A. C., MC NAMARA, D. S. & LOUWERSE, M. M. (2003). « What do readers need to learn in order to process coherence relations in narrative and expository text.» *In* A. P. Sweet & C.E. Snow (eds), *Rethinking reading comprehension*, New-York: Guilford Press, 82-98.

GRANGER, S. (2009). Learner corpora. *In* Lüdeling, A. & Kytö, M. (eds.), *Corpus linguistics. An international handbook.* Vol 1, Berlin/New-York: de Gruyter, 259-274.

GRESILLON, A. (1994). *Éléments de critique génétique : lire les manuscrits modernes*. Paris : Presses Universitaires de France.

GROSZ, B.J., JOSHI, A.K., WEINSTEIN, S. (1995). Centerring: a framework for modelling the local coherence of discourse. *Language*, n°69-2, 274-307.

GUILLOT, C. (2006). Démonstratif et deixis discursive : analyse compare d'un corpus écrit en français médiéval et d'un corpus oral de français contemporain. *Langue française*, n°152, 56-69.

GUNDEL, J., HEDBERG, N. & ZACHARSKI, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, n°69 (2), 274-307.

GUNDEL, J., HEDBERG, N. & ZACHARSKI, R. (2000). Statut cognitif et forme des anaphoriques indirects. *Verbum*, XXII, 1, 79-102.

GUNNARSSON-LARGY, C. & AURIAC-SLUSARCZYK, E. (2013). *Écriture et réécritures chez les élèves. Un seul corpus, divers genres discursifs et méthodologies d'analyse*. Sciences du langage : Carrefours et points de vue n°10, Louvain-La-Neuve : Academia L'Harmattan.

HABERT, B. (2005). Portrait de linguiste(s) à l'instrument. *Texto !* vol. X, n°4, [En ligne].

HABERT, B., NAZARENKO, A., & SALEM, A. (1997). *Les linguistiques de corpus*. Paris : Armand Colin.

HALLIDAY, M.A.K. (1967). Notes on transitivity and theme in English, Part 2. *Journal of Linguistics*, n°3-2, 199-244.

HALLIDAY, M.A.K. & HASAN, R. (1976). *Cohesion in English*. London: Longman.

HALTE, J-F. (1982). Écrire un roman historique en classe de cinquième. *Repères,* n°36, 49-58.

HALTE, J-F. (1989). Analyse de l'exercice dit « la rédaction » et propositions pour une autre pédagogie de l'écriture. *In* Charolles, M., Halté, J-F., Masseron, C., & Petitjean, A. *Pour une didactique de l'écriture*. Metz : Centre d'analyse syntaxique de l'université de Metz, 9-47.

HAWKINS, J.A. (1977a). The pragmatics of definiteness. Part I. *Linguistische Berichte*, n°47, 1-27.

HAWKINS, J.A. (1977b). The pragmatics of definiteness. Part II. *Linguistische Berichte*, n°48, 1-27.

HAWKINS, J.A. (1978). *Definiteness and indefiniteness; a study in reference and grammaticality prediction*. London: Croom Helm.

HAYES, J.R. (1996). A new framework for understanding cognition and affect in writing. *In* Levy, C. & Ransdell, S. (eds.), *The science of writing: theories, methods, individual differences and applications,* Mahwah: Lawrence Erlbaum Associates, 1-28.

HAYES, J-R. & FLOWER, L. (1980). Identifying the organization of writing processes; the dynamic of composing: making plans and juggling constraints. *In* Gregg, W. & Steinberg, E. *Cognitive processes in writing.* Hildale (NJ): Lawrence Erlbaum Associates Publishers.

HAYES, J.R. & FLOWER, L. (1981). Uncovering cognitive processes in writing: an introduction to protocol analysis. *Research on writing*, 207-220.

HEURLEY, L. (2006). La révision de texte : l'approche de la psychologie cognitive. *Langages,* n°164, 10-25.

HICKMANN, M. (1980). Creating referent in discourse: a developmental analysis of linguistic cohesion. *In* Kreiman, J. & Ojeda, A.E. (eds.), *Papers from the parasession on pronouns and anaphora*, Chicago: Chicago Linguistic Society, 192-203.

HICKMANN, M. (1984). L'ontogenèse de la cohésion dans le discours. *In* Deleau, M. (ed.), *Langage et communication à l'âge préscolaire*, Rennes : Presses Universitaires de Rennes 2, 27-57

HICKMANN, M. (1987). Ontogenèse de la cohésion dans le discours. *In* Piéraut-Le Bonniec, G. (Ed.), *Connaître et le dire*, 239-262. Bruxelles : Mardaga.

HICKMANN, M. (1995). Discourse organization and the development of reference to person, space and time. In Fletcher, P., MacWhinney, B. (eds.), *The Handbook of child language*, Oxford: Blackwell, 194-218.

HICKMANN, M. (2000). Le développement de l'organisation discursive. *In* Kail, M. & Fayol, M. (eds.), *L'acquisition du langage. Le langage en développement au-delà de trois ans*, Paris : PUF, 83-116.

HICKMANN, M. (2003). *Children's discourse: person, space and time across languages.* Cambridge ; New-York: Cambridge University Press.

HICKMANN, M. (2004). Le développement de la cohésion dans la narration orale chez l'enfant : perspectives inter-langues. CALAP, n°24, 13-31.

HICKMANN, M. & SCHNEIDER, P. (1993). Children's ability to restore the referential cohesion of stories. *First language*, n°13, 169-202.

HICKMANN, M., KAIL, M., ROLAND, F. (1995). Organisation référentielle dans les récits d'enfants en fonction des contraintes contextuelles. *In Enfance*, tome 48, n°2. Aspects actuels de recherches sur la Psychologie du Développement, 215-229.

HIMMELMANN, N. (1996). Demonstratives in narrative discourse: a taxonomy of universal uses. *In* Fox, B. (eds.), *Studies in Anaphora*, Amsterdam/philadelphia: John Benjamins Publishing, 202-254.

HIRSCHMANN, H. & *al.* (2013). Underuse of syntactic categories. A case study on modification. *In* Granger, S., Gilquin, G. & Meunier, F. (Hrsg.), *20 years of learner corpus research. Looking back, moving ahead. Corpora and language in use.* Proceedings 1. Louvain la Neuve : Presses Universitaires de Namur.

HUXLEY, R. (1970). The development of the correct use of subject personal pronouns in two children. *In* Florès d'Arcais, G.B. & Levelt, W.J.M., *Advances in Psycholinguistics*, Amsterdam-London: North Holland Publishing Co, 141-165.

JACQUES, M-P. & RINCK, F. (2017). Un "corpus de littéracie avancée" : résultat et point de départ. *Corpus*, n°16, 217-237.

JAFFRE, J-P. & FAYOL, M. (1989). Structurations de textes : connecteurs et démarcation graphiques. *Langue Française*, n°81.

JAUBERT, A. (2005). *Cohésion et cohérence. Études de linguistique textuelle.* Sous la direction d'Anna Jaubert, Lyon : ENS Éditions.

JISA, H. (2000). Increasing cohesion in narratives: a developmental study of maintaining and reintroducing subjects in French. *Linguistics*, n°38-3, 591-620.

JISA, H. (2004). Augmenter la cohésion dans des textes narratifs : une étude développementale du maintien et de la réintroduction des sujets grammaticaux. *CALAP*, n°24, 33-56.

JOHNSEN, L.A. (2014). Parenthèses ana-cataphoriques et figures du discours. *In* Gramm-R Etudes de linguistique française, n°27 (eds.), *Linguistique et stylistique des figures*, 123-146.

JOHNSON-LAIRD, P. (1980). Mental models in cognitive science. *Cognitive Science*, n°4, 71-115.

JOHNSON-LAIRD, P. (1983). *Mental models.* Cambridge: Cambridge University Press.

KAIL, M. & LEVEILLE, M. (1977). Compréhension de la coréférence des pronoms personnels chez l'enfant et l'adulte. *Année Psychologique*, n°77, 79-94.

KAIL, M. & HICKMANN, M. (1992). French children's ability to introduce referents in narratives as a function of mutual knowledge, *First Language*, n°12, 637-662.

KAIL, M. & SANCHEZ Y LOPEZ, I. (1997). Referent introduction in Spanish narratives as a function of contextual constraints: acrosslinguistic perspective, *First language*, n°17, 103-130.

KARA, M. & WIEDERSPIEL, B. (2011). Anaphore résomptive conceptuelle et mémoire discursive : entre identité et altérité. *Itinéraires*, 2011-2, 79-93.

KARMILOFF-SMITH, A. (1979). *A functional approach to child language: A study of determiners and reference.* Cambridge: Cambridge University Press.

KARMILOFF-SMITH, A. (1980). Psychological processes underlying pronominalisation and non-pronominalisation in children's connected discourse. *In* Kreiman, J. & Ojeda, A.E. (eds.), *Papers from the Parasession on Pronouns and Anaphora.* Chicago: Chicago Linguistic Society.

KARMILOFF-SMITH, A. (1981). The grammatical marking of thematic structure in the development of language production. *In* Deutsch, W. (ed.), *The Child's Construction of Language.* London: Academic Press.

KARMILOFF-SMITH, A. (1983). Language Development as a problem-solving process. *Papers and Reports on Child Language Development,* n°22, 1-22.

KELLOG, R.T. (1987). Effects of topic knowledge on the allocation of processing time and cognitive effort to writing processes. *Memory and Cognition*, n°15, 256-266.

KESIK, M. (1989). *La cataphore.* Presses Universitaire de France.

KLEIBER, G. (1981). *Problèmes de référence : descriptions définies et noms propres.* Paris : Klincksieck.

KLEIBER, G. (1988). Peut-on définir une catégorie générale de l'anaphore ? *Vox Romanica,* n° 47, [En ligne].

KLEIBER, G. (1990). Marqueurs référentiels et processus interprétatifs : pour une approche « plus sémantique ». *Cahiers de linguistique française,* n°11, 241-258.

KLEIBER, G. (1991). Anaphore-deixis : où en sommes-nous ? *L'information Grammaticale,* n°51, 3-18.

KLEIBER, G. (1992). Article défini, unicité, pertinence. *Revue romane*, n°27/1, 61-89.

KLEIBER, G. (1994). *Anaphores et pronoms*. Louvain La Neuve : Duculot.

KLEIBER, G. (2001). Regards sur l'anaphore et la givenness hierarchy. *In Langage et Référence, Acta Universitatis*, 311-322.

KHOROUNJAIA, E. & TOLCHINSKY, L. (2004). Discursive constraints on the lexical realization of arguments in Spanish. *In* Berman, R. (eds.), *Language Development across Childhood and Adolescence*, Amsterdam/Philadelphia: John Benjamins.

LAMBRECHT, K. (1994). *Information structure and sentence form: topic, focus and mental representations of discourse referents.* Cambridge : Cambridge University Press.

LAMBRECHT, K. (2000). Prédication seconde et structure informationnelle : la relative de perception comme construction présentative. *Langue française,* n°127, 49-66.

LANDRAGIN, F. (2004). Saillance physique et saillance cognitive. *Corela,* n°2-2, [En ligne].

LANDRAGIN, F. (2010). Sur les aspects multicritères et multidimensionnels de la saillance. 5e Colloque du Réseau des linguistes du Grand Est, novembre 2010, Strasbourg, France., 30-33.

LANDRAGIN, F. (2011). Une procédure d'analyse et d'annotation des chaînes de coréférence dans des textes écrits. *Corpus,* n°10, [En ligne].

LANDRAGIN, F. (2014). Anaphores et coréférences : analyse assistée par ordinateur. In M. Fossard & M-J Béguelin (éds.). *Nouvelles perspectives sur l'anaphore.* Berne : Peter Lang, 29-54.

LANDRAGIN, F., POIBEAU, T., VICTORRI, B. (2012). ANALEC : a New Tool for the Dynamic Annotation of Textual Data. European Language Resources Association (ELRA). *International Conference on Language Resources and Evaluation (LREC 2012), mai 2012*, Istanbul, Turquie.

LANDRAGIN, F., TANGUY, N., CHAROLLES, M. (2015). Références aux personnages dans *L'Occupation des sols* : apport de la linguistique outillée. *Revue Sciences/Lettres*, [En ligne], n°3, 1-21.

LEECH, G. (1997). Introducing corpus annotation. *In* Garside, R., Leech, G. & Mcenery, A. (eds.), *Corpus annotation: linguistic information from computer text corpora.* London: Longman, 1-18.

LENTZ, F. (1993). Travailler la cohésion textuelle : quelques exemples pratiques. *Québec français*, n°88, 58-65.

LEON, J. (2008). Aux sources de la « Corpus Linguistics » : Firth et la *London School Langages*, n°171, 12-33.

LEVELT, W. (1989). *Speaking: from intention to articulation.* MIT Press.

LONGO, L. (2013). *Vers des moteurs de recherche « intelligents » : un outil de détection automatique de thèmes. Méthode basée sur l'identification automatique des chaînes de référence.* Thèse de doctorat de l'université de Strasbourg.

LONGO, L. & TODIRASCU, A. (2009). Une étude de corpus pour la détection automatique de thèmes. *Textes et Corpus*, n°4, 143-155.

LOPEZ, S. (2010). Corpus de référence et corpus d'usages : méthodologie de constitution pour une analyse des communications pilote-contrôleur. *Les cahiers de praxématique*, Montpellier : Presses Universitaires de la Méditerranée, 2006-2010, 59-72.

LUDELIN, A. (2008). Mehrdeutigkeiten und Kategorisierung : Probleme bei der Annotation von Lernerkorpora. *In* Walter, M. & Grommes, P. (Hrsg.), *Fortgeschrittene Lernervarietäten*. Tübingen : Niemeyer, 119-140.

LUNDQUIST, L. (1980). *La cohérence textuelle. Syntaxe, sémantique, pragmatique.* Copenhague : Nyt Nordisk Forlag.

LYONS, J. (1980). *Sémantique linguistique*. Paris : Larousse.

MAILLARD, M. (1974). Essai de typologie des substituts diaphoriques. *Langue française*, n°21, 55-71.

MARANDIN, J-M. (1986). *Ce* est un autre. L'interprétation anaphorique du syntagme démonstratif. *Langage*, n°81, 75-89.

MARIN, B. (2007). Culture écrite, culture orale : des ressources spécifiques. *La linguistique*, n°43, 105-122.

MARMY CUSIN, V. (2012). *Développer et comprendre des pratiques d'enseignement de la grammaire intégrées à la production textuelle : entre les dires et les faires.* Thèse de doctorat de l'Université de Genève.

MAS, M. (1989). Aspects du traitement didactique des référents. « Embarquement pour Critères ». *Repères pour la rénovation de l'enseignement du français*, n°7, 7-21.

MAS, M., GARCIA-DEBANC, C., ROMIAN, H., SEGUY, A., TAUVERON, C. et TURCO, G. (1991). *Comment les maîtres évaluent-ils les écrits de leurs élèves en classe ?* Paris : INRP-Hachette Éducation.

MASSERON, C. (1989). La correction de rédaction. *In* Charolles, M., Halté, J-F., Masseron, C. & Petitjean, A. *Pour une didactique de l'écriture*. Metz : Centre d'analyse syntaxique de l'université de Metz, 85-106.

MASSERON, C. (2005). Indicateurs langagiers et stratégies scripturales. Du discours à la langue. *Pratiques*, n°125-126, 205-249.

MASSERON, C. (2011). L'analyse linguistique des écrits scolaires. *Pratiques*, n°149-150, 129-162.

MASSERON, C. (2014). Contradictions et nécessités de l'enseignement de la grammaire : la difficulté empirique des énoncés complexes en production, entre morphosyntaxe, lexique et pragmatique. *Repères*, n°50, [En ligne].

MASSERON, C. (2016). Présentation. *Pratiques*, n°169-170 [En ligne].

MASSERON, C. & SCHNEDECKER, C. (1995). Production des chaînes de référence et problèmes de structuration textuelle : la redénomination comme catalyseur d'une conjoncture discursive problématique. *Linx,* H-S n°6-2, 423-442.

MAZUR-PALANDRE, A. (2009). *Le flux de l'information, aspects syntaxiques et discursifs, une étude fonctionnaliste et développementale.* Thèse de Doctorat de l'université de Lyon 2, soutenue le 18 septembre 2009.

MELANIE-BECQUET, F. & LANDRAGIN, F. (2014). Linguistique outillée pour l'étude des chaînes de référence : questions méthodologiques et solutions techniques. *Langages,* n°195, 117-137.

MILLOGO, V.E. (2008). Gestion des expressions anaphoriques en production de textes chez des enfants de 7 à 12 ans : effets de quelques variables contextuelles. *In* Alamargot, D., Bouchand, J., Lambert, E., Millogo, V. & Beaudet, C. (eds.), *Proceedings of the International Conference « de la France au Québec : l'Écriture dans tous ses états »*, Poitiers, France, 12-15 novembre 2008.

MILNER, J-C. (1976). Réflexions sur la référence. *Langue française*, n°30, 63-73.

MILNER, J-C. (1982). *Ordres et raisons de langue.* Paris : Éditions du Seuil.

MOIRAND, S. (1975). Le rôle anaphorique de la nominalisation dans la presse écrite. *Langue Française,* n°28, 60-78.

MORTUREUX, M-F. (1998). *La lexicologie entre langue et discours.* Paris : Armand Colin.

NONNON, E. (2010). La notion de progression au cœur des tensions de l'activité d'enseignement. *Repères*, n°41, 5-34. [En ligne].

PAOLACCI, V. & FAVART, M. (2010). Traitement des marques de cohésion par les jeunes scripteurs : l'utilisation de la ponctuation et des connecteurs à l'entrée en sixième. Approche linguistique, cognitive et didactique. *Langages,* n°177, 113-128.

PAOLACCI, V. & ROSSI-GENSANE, N. (2012). Quelles images de la phrase dans les écrits d'élèves de fin d'école primaire française ? *Congrès Mondial de Linguistique Française, CMLF 14, juillet 2014*, Berlin, Allemagne.

PARET, M-C. (2003). La « grammaire » textuelle : une ressource pour la compréhension et l'écriture des textes. *Québec français,* n°128, 48-50.

PATTABHIRAMAN, T. (1993). *Aspects of salience in natural language generation.* PH.D. Thesis, Simon Fraser University.

PAVEAU, M-A. (2012). Genre de discours et technologie discursive. Tweet, twittécriture et twittérature. [En ligne].

PELLEGRINI, A.D., GALDA, L., RUBIN, D.L. (1984). Context in text: The development of oral and written language in two genres. *Child Language*, n°3, 63-75.

PETITJEAN, A. (1999). Textes officiels et enseignement du français. *Pratiques*, n°101/102, 3-7.

PLANE, S. (1995). Pratiques sociales expertes et dispositifs didactiques d'écriture sur traitement de textes. *Repères,* n°11, 103-123.

PLANE, S. (2000). Éléments pour un usage didactique du traitement de texte. Écrire, réécrire et réviser sur ordinateur. *Pratiques*, n°105-6, 159-181.

POESIO, M. & VIEIRA, R. (1998). A corpus-based investigation of definite description use. *Journal Computational Linguistics*, vol.24-2, 183-216.

PRINCE, E.F. (1981). Toward a taxonomy of Given-New information. *In* Cole, P. (eds.), *Radical Pragmatics.* New_York: Academic Press, 223-255.

RAGAN, P.H. (2001). Classromm use of a systemic functional small learner corpus. *In* Ghadessy, A., Henry, A., Roseberry, R.L. (eds.), *Small corpus studies and ELT. Theory and practice.* Amsterdam/Philadelphia: Benjamins, 207-236.

RASTIER, F. (2001). *Arts et sciences du texte.* Paris : PUF.

REBOUL, A. (1989). Résolution de l'anaphore pronominale sémantique et/ou pragmatique. *Cahiers de linguistique française*, n°10, 77-100.

REICHLER-BEGUELIN, M.J. (1988a). Anaphore, cataphore et mémoire discursive. *Pratiques,* n°57, 15- 40.

REICHLER-BEGUELIN, M.J. (1988b). Norme et textualité. Les procédés référentiels considérés comme déviants en langue écrite. *In* Schoeni, G., Bronckart, J-P. & Perrenoud, P. (eds.), *La langue française est-elle gouvernable ? Normes et activités langagières*, Neuchâtel et Paris : Delachaux et Niestlé, 185-216.

REICHLER-BEGUELIN, M.J. (1989). « Anaphores, connecteurs et processus inférentiels ». In C. Rubattel (éd.), *Modèles du discours. Recherches actuelles en Suisse Romande, Berne : Peter Lang, 303-336.*

REICHLER-BEGUELIN, M-J. (1993). L'encodage du texte écrit. Normes et déviances dans les processus référentiels et dans le marquage de la cohésion. European Science Foundation Network on Written Language and Literacy, second Workshop « Understanding early literacy in a developmental and cross-linguistic approach », octobre 1993, Wassenaar, Pays-Bas.

REICHLER-BEGUELIN, M-J. (1995). Déficits dans la maîtrise des procédés de cohésion. *Linx*, hors-série n°6-2, 405-422.

REICHLER-BEGUELIN, M-J. (1999). « De la phrase aux énoncés » : bilan et perspectives. *Tranel,* n°31, 143-153.

REINHART, T. (1980). Conditions for Text Coherence. *Poetics Today*, vol. 1, n°4, 161-180.

RIEGEL, M., PELLAT, J6C., RIOUL, R. (1994). *Grammaire méthodique du français.* Paris : PUF.

RONDELLI, F. (2008). La cohérence des textes : entre préoccupations didactiques et questionnements linguistiques. Actes du Congrès Mondial de Linguistique Française (CMLF'08), Durand J., Habert B., Laks B. (éds.).

RONDELLI, F. (2010a). La cohérence textuelle : pratiques des enseignants et théories de référence. *Pratiques*, n°145/146, 55-84.

RONDELLI, F. (2010b). Comment les enseignants construisent-ils un objet de savoir ? Exemple de la cohérence textuelle. *Repères,* n°42, 63-81.

ROUBAUD, M-N. (2015). Accéder aux compétences syntaxiques des élèves (8-11 ans) : le cas des relatives. *In* Trévisiol-Okamura, P. & Kaheraoui, M. *Les subordonnées. Corpus, acquisition et didactique*, Rennes : Presses Universitaires de Rennes, 123-138.

ROUBAUD, M-N. (2017). Le français écrit : transcription et édition. Le cas des textes scolaires. *Corpus*, n°16, 113-131.

ROUSSEY, J-Y. & PIOLAT, A. (2005). La révision du texte : une activité de contrôle et de réflexion. *Psychologie Française*, n°50, 351-372.

RUTTER, P. & RABAN, B. (1982). The development of cohesion in children's writing: A preliminary investigation. *First Language*, n°3, 63-75.

SALAZAR-ORVIG, A., FAYOLLE, V., HASSAN, R., LEBER-MARIN, J., MARCOS, H., MORGENSTERN, A., PARES, J. (2004). Émergence des marques anaphoriques : le cas des pronoms. *CALAP*, n° 24, 57-82.

SANFORD, A. & GARROD, S. (1981a). *Understanding written language*. New-York: Wiley.

SANFORD, A. & GARROD, S. (1981b). Vers la construction d'un modèle psychologique de la compréhension du langage écrit. *Bulletin de psychologie,* n°35, 643-648.

SCARDAMALIA, M. & BEREITER, C. (1983). The development of evaluative, diagnostic and remedial capabilities in children's composing. *In* Martlew, M. (ed.), *The psychology of written language: a developmental approach*, London: Wiley, 67-95.

SCHNEDECKER, C. (1995). Besoins didactiques en matière de cohésion textuelle : les problèmes de continuité référentielle. *Pratiques,* n°85, 3-25.

SCHNEDECKER, C. (1997). *Nom propre et chaînes de référence*. Metz : Centre d'Études Linguistiques des Textes et Discours.

SCHNEDECKER, C. (2005). Les chaînes de référence dans les portraits journalistiques : éléments de description. *Travaux de linguistique,* n°51, 85-133.

SCHNEDECKER, C. (2009). La notion de saillance : problèmes définitoires et avatars. *Aspects linguistiques et communicatifs de la mise en évidence dans un texte*, colloque Saillance, Université de Genève, 3-6.

SCHNEDECKER, C. & CHAROLLES, M. (1993). Coréférence et identité : le problème des référents évolutifs. *Langages,* n°112, 106-126.

SCHNEDECKER, C. & LONGO, L. (2012). Impact des genres sur la composition des chaînes de référence : le cas des faits divers. Congrès Mondial de la Linguistique française, CMLF 2012, juillet 2012, Lyon, France.

SCHNEDECKER, C. & LANDRAGIN, F. (2014). Les chaînes de référence : présentation. *Langages,* n°195, 3-22.

SCHNEDECKER, C., GLIKMAN, J. & LANDRAGIN, F. (2017). Les chaînes de référence : annotation, application et questions théoriques. *Langue française*, n°195, 5-16.

SCHNEUWLY, B. (1988). *Le langage écrit chez l'enfant*. Neuchâtel : Delachaux et Niestlé.

SIMARD, C., DUFAYS, J-L., DOLZ, J. & GARCIA-DEBANC, C. (2010). *Didactique du français langue première.* Bruxelles : De Boeck.

SIMONS, (1966). La ponctuation dans l'expression écrite chez l'enfant. *Bulletin de psychologie,* n°29, 525-535.

SINCLAIR, J. (1996). *Preliminary recommendations on Corpus Typology,* Rap. Tech. EAGLES (Expert Advisory Group on Language Engineering Standards), may 19996, CEE.

SPERBER, D. & WILSON, D. (1989). *La pertinence*. Paris : Éditions de Minuit.

TANZ, C. (1974). Cognitive principles underlying children's errors in pronominal case-marking, *Journal of Child Language*, n°1, 271-276.

TAUVERON, C. (1987). "On ne sait pas de qui tu parles". Un problème de mise en texte : comment reprendre et distinguer les personnages d'un récit. Exemple de traitement au cours préparatoire. *Repères*, n°73, 27-35.

TAUVERON, C. (1995). Le personnage, articulateur privilégié (mais problématique) de lecture-écriture à l'école élémentaire. *Pratiques*, n°86, 27-49.

TESNIERE, L. (1959). *Éléments de syntaxe structurale.* Paris : Klincksieck.

TESTENOIRE, P-Y. (2017). Transcrire des écrits scolaires : entre philologie et génétique textuelle. *Corpus*, n°16, 87-109.

TRAXLER, M. & GERNSBACHER, M. (1995). Improving coherence in written communication. *In* Gernsbacher, M., Givón, T. (eds.), *Coherence in spontaneous text*, Philadelphia: John Benjamins, 215-238.

TUTIN, A., CLOUZOT, C., ANTONIADIS, G. (2000a). Un corpus d'anaphores discursives pour les études en TAL. Rapport non publié, [En ligne], http://w3.u-grenoble3.fr/tutin/Publis/rapport.pdf.

TUTIN, A., TROUILLEUX, F., CLOUZOT, C. GAUSSIER, E., ZAENEN, A., RAYOT, S., ANTONIADIS, G. (2000b). Annotating a large corpus with anaphoric links, *Proceedings of DAARC 2000.*

TUTIN, A. & GROSSMAN, F. (2002). Collocations régulières et irrégulières : esquisse de typologie du phénomène collocatif. *Revue française de linguistique appliquée,* vol. VII, 7-25.

UMSTEAD, R.S. & LEONARD, L.B. (1983). Children's resolution of pronominal reference in text, *First language*, n°4, 73-84.

VAN DE VELDE, D. (1997). Articles, généralités, abstractions. *In* Flaux, N., Van de Velde, D., & De Mulder, W. (eds), *Entre général et particulier : les déterminants*, Arras : Artois Presses Universitaires, 86-136.

VAN DIJK, T. & KINTSCH, W. (1983). *Strategies of discourse comprehension*. New-York: Academic Press.

WANNER, L. & BATEMAN, J.A. (1990). A collocational based approach to Salience-Sensitive lexical selection. *In* Proceedings of the fifth international workshop on natural language generation, Dawson, 31-38.

WARDEN, D.A. (1976). The influence of context on children's use of identifying expressions and reference. *British Journal of Psychology,* n°67, 101-112.

WARDEN, D.A. (1981). Learning to identify referents. *British Journal of Psychology,* 67(1), 101-112.

WARYAS, C.L. (1973). Psycholinguistic research in language intervention programming: the pronoun system, *Journal of Psycholinguistic Research*, n°3, 221-235.

WIDLOCHER, A. & MATHET, Y. (2009). La plate-forme Glozz : environnement d'annotation et d'exploration de corpus. TALN 09, Conférence sur le Traitement Automatique des Langues Naturelles, juin 2009, Senlis, France.

WIGGLESWORTH, G. (1990). Children's narrative acquisition: a study of some aspects of reference and anaphora. *First Language*, n°10, 105-125.

WILMET, M. (1986). *La détermination nominale*. Paris : PUF.

WOLFARTH, C., PONTON, C. & TOTEREAU, C. (2017). Apports du TAL à la constitution et à l'exploitation d'un corpus scolaire au travers du développement d'un outil d'annotation orthographique. *Corpus,* n°16, 185-214.

YULE, G. (1982). Interpreting anaphora without identifying reference. *Journal of Semantics,* n°1, 315-323.

Textes concernant les programmes d'enseignement

Plan de rénovation de l'enseignement du français, *L'enseignement public*, février 1971.

Instructions relatives à l'enseignement du français à l'école élémentaire, 1972. Circulaire n° 72-474 du 4 décembre 1972.

Terminologie grammaticale pour l'école élémentaire, 1976. Circulaire n°76-363 du 25 octobre 1976.

Horaires, objectifs et programmes pour le Cycle Élémentaire, 1978. Arrêté du 7 juillet 1978.

Horaires, objectifs et programmes pour le Cycle Moyen, 1980. Arrêté du 18 juillet 1980.

Horaires, Programmes et Instructions pour l'école élémentaire, 1985. Arrêté du 15 mai 1985.

Collèges, Programmes et Instructions, BO n°44 du 12-12-1985.

Les cycles à l'école primaire, 1991.

La charte des programmes, 1992.

La Maîtrise de la Langue à l'École, 1992.

Programme de l'école primaire, 1995. Arrêté du 22-2-1995, JO du 2-3-1995.

Programme du collège, 1995. Arrêté du 22 novembre 1995, BO n°48 du 28 décembre 1995.

Horaires et Programmes d'Enseignement de l'École Primaire, 2002. Arrêté du 25-1-2002, JO du 10-2-2002.

Lire et écrire au cycle 3. Repères pour organiser les apprentissages au long du cycle. (2003). Document d'accompagnement des programmes, CNDP.

Les nouveaux programmes de l'école primaire, 2008. BO 20-2-2008, n°0 H-S.

Programmes du collège. Programmes de l'enseignement de français, 2008. BO Spécial du 28-08-2008, n°6.

Programmes d'enseignement de l'école élémentaire et du collège, 2015. BO spécial du 26-11-2015, n°11.