



**HAL**  
open science

## **La continuité référentielle dans le corpus RÉSOLCO : méthode d'annotation et premières analyses**

Lydia-Mai Ho-Dac, Claudine Garcia-Debanc, Josette Rebeyrolle

### ► **To cite this version:**

Lydia-Mai Ho-Dac, Claudine Garcia-Debanc, Josette Rebeyrolle. La continuité référentielle dans le corpus RÉSOLCO : méthode d'annotation et premières analyses. Langue française, 2021, 211 (3), pp.99-114. <10.3917/lf.211.0099>. <hal-03559961>

**HAL Id: hal-03559961**

**<https://hal.science/hal-03559961v1>**

Submitted on 11 Feb 2025

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.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

# La continuité référentielle dans le corpus Résolco : méthode d'annotation et premières analyses

## Referential continuity ties in the Resolco Corpus: annotation and first results

Claudine Garcia-Debanco  
CLLE, Université de Toulouse, CNRS

Josette Rebeyrolle  
CLLE, Université de Toulouse, CNRS

Lydia-Mai Ho-Dac  
CLLE, Université de Toulouse, CNRS

### Résumé

Le corpus Résolco est constitué de 400 textes rédigés par des élèves français âgés de 9 à 15 ans en réponse à une consigne d'écriture conçue comme une tâche impliquant la résolution de problèmes de cohésion. Cette contribution expose une méthode d'annotation de la continuité référentielle dans la désignation des personnages principaux d'un texte narratif et l'illustre avec l'analyse de trois textes d'élèves. Les configurations de marques linguistiques observées sont interprétées comme indices de la stratégie des rédacteurs. Sont présentés les premiers résultats de l'analyse d'une centaine de textes de trois niveaux contrastés (CE2, 6e et 3e).

### Mots-clés

cohésion textuelle, corpus, textes scolaires, annotation discursive, stratégies rédactionnelles

### Abstract

The ResolCo Corpus is composed of 400 texts written by French students aged 9-15 in response to a task related to "Resolving Cohesion problems". This paper proposes a method for describing the strategies used by writers for creating cohesive ties and more specifically referential continuity ties. This method involves systematic annotation of the linguistic expressions referring to the main characters that must occur in all texts given the instruction. An analysis of a first set of annotations across three education levels outlines two prototypical and opposite strategies: a step by step strategy vs. a planning strategy.

### Keywords

cohesion ties, reference, discourse annotation, student's writing, discourse strategies

## 1. INTRODUCTION

La mise en place d'un enseignement de la cohésion textuelle suppose une connaissance précise des formes linguistiques utilisées par des rédacteurs experts pour assurer la cohérence et la cohésion des textes ainsi que des informations sur leur acquisition. Une cartographie de l'emploi de ces formes linguistiques (telles que les procédés de coréférence, les temps verbaux, la ponctuation, la segmentation en paragraphes, les connecteurs, les organisateurs textuels) en fonction du niveau scolaire et du degré d'expertise du rédacteur doit être réalisée si l'on veut déterminer des indicateurs de progression dans le domaine de l'enseignement de la textualisation<sup>1</sup> fondés scientifiquement, comme souhaité par Nonnon (2010). L'élaboration de cette cartographie suppose d'inventorier les formes linguistiques majoritairement utilisées par les élèves d'un niveau scolaire donné, et, pour ce faire, de procéder au traitement exhaustif et systématique d'un grand nombre de données, comme cela a pu être fait pour l'orthographe (Geoffre, 2014 ; Brissaud, Viriot-Goeldel, 2019 ; ici même David, Brissaud).

Notre projet est de réaliser un travail analogue dans le domaine de la cohésion textuelle et plus précisément du point de vue de la construction de la référence. Dans le présent article, nous présentons une méthode dont l'objectif est double : (a) constituer une ressource annotée en termes de continuité référentielle composée de textes d'élèves de niveaux scolaires variés et (b) élaborer une cartographie des formes linguistiques manifestant les compétences textuelles et discursives en cours de construction du milieu de l'école primaire à la fin du collège<sup>2</sup>. Après avoir exposé le contexte de l'étude (section 2), nous présenterons les choix méthodologiques que nous avons retenus pour l'annotation des continuités référentielles (section 3), puis nous exposerons une première esquisse des stratégies utilisées par les élèves pour construire des continuités référentielles (section 4). La dernière section (section 5) sera consacrée à la présentation des premiers résultats obtenus sur une centaine de textes.

## 2. CONTEXTE DE L'ÉTUDE

Notre projet de décrire de façon exhaustive et systématique les formes linguistiques utilisées pour construire la référence dans des textes d'élèves s'inscrit dans la tâche Cohérence du projet É-Calm<sup>3</sup>. Les textes analysés dans notre étude

---

1 Le terme de « processus de textualisation » est proposé par Cislaru & Olive (2018) pour rendre compte d'un nouveau chantier de recherche d'analyse d'écritures en temps réel « en jetant des ponts entre les préoccupations de la linguistique des textes et des usages d'une part et de la psycholinguistique de l'écrit d'autre part, grâce aux possibilités technologiques offertes par les logiciels de suivi de rédaction qui enregistrent le processus d'écriture » (Cislaru & Olive, 2018, 9). Les auteurs ont travaillé sur des enregistrements d'écritures en temps réel réalisées par des adultes sur des genres codifiés (rapports pour les services sociaux et dossiers universitaires). Bien que nos propres travaux n'utilisent pas les méthodologies d'écriture *on line*, ils essaient d'inférer à partir des marques linguistiques la dynamique de l'écriture et les stratégies des rédacteurs, en confrontant les résultats à ceux des études psycholinguistiques.

2 Une première cartographie de ces formes linguistiques à partir de l'analyse de 210 textes d'école primaire en réponse à cette même consigne a été réalisée en 2018 par Karine PEREZ-BONNEMAISON dans sa thèse, *Anaphore et référence en production écrite : étude de textes narratifs d'élèves de 9 à 11 ans, du CE2 au CM2*, soutenue à l'Université Toulouse Jean-Jaurès le 14 juin 2018.

3 L'ANR *É-Calm Écriture scolaire et universitaire : Corpus, Analyses Linguistiques, Modélisations Didactiques*, projet ANR-17-CE28-0004 du programme Sociétés innovantes, intégrantes et

correspondent aux premières versions de productions écrites réalisées dans le cadre habituel de la classe en réponse à une même consigne. 400 textes ont été collectés dans des classes de fin d'école primaire et de collège (du CE2, 3<sup>e</sup> année de l'école primaire, élèves de 9 ans, à la troisième, élèves de 15 ans). Ces textes constituent le corpus Résolco.

## 2.1. Le corpus Résolco

La tâche d'écriture choisie peut être considérée comme une tâche de résolution de problèmes de cohésion textuelle, d'où le nom de Résolco (Résolution de Problèmes de Cohésion textuelle). La consigne est la suivante<sup>4</sup> :

*Raconte une histoire dans laquelle tu insèreras séparément et dans l'ordre donné 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.*

Les trois phrases de la consigne sont inscrites sur des bandelettes de papier que l'élève peut coller quand il le souhaite. Ce dispositif vise à limiter les altérations portées sur les phrases à insérer et à ne pas induire une longueur de texte à rédiger. Chacune de ces phrases comporte des anaphores : pronoms personnels de troisième personne (*elle et il*) et syntagmes nominaux démonstratifs (*cette maison, ce grand bruit, cette aventure*)<sup>5</sup>.

La tâche d'écriture impose l'insertion de ces phrases dans un texte narratif fictionnel. Elle fait écho aux nombreuses difficultés rencontrées par les élèves d'école primaire et de collège dans la gestion des procédés de reprise lorsqu'ils rédigent des textes (Schnedecker, 1995 ; Masseron, 2005). Contrairement aux consignes d'écriture habituelles, dans lesquelles le choix de l'identité référentielle du personnage est réalisé antérieurement à la mise en texte<sup>6</sup>, la consigne Résolco impose d'instancier référentiellement un personnage à partir d'une expression anaphorique imposée.

## 2.2. Le système récit-personnage dans les travaux en didactique du français langue première

Depuis le travail fondateur de Tauveron (1995), le personnage est considéré comme une « clef pour la didactique du récit à l'école élémentaire » aussi bien pour

---

adaptatives (DS08) 2017, Janvier 2018-Juin 2022, coordonné par Claire Doquet, associée des enseignants-chercheurs de quatre équipes de recherche : CLESTHIA, LIDILEM, CLLE et CIRCEFT.

<sup>4</sup> On pourra lire des éléments complémentaires sur cette consigne dans des publications antérieures (Garcia-Debanc, Bonnemaïson, 2014 ; Garcia-Debanc *et al.*, 2017 ; Garcia-Debanc, 2020). Bien que le temps d'écriture n'ait pas été limité, il ne dépasse généralement pas une trentaine de minutes étant donné que les textes récoltés ont été produits en classe et n'ont pas donné lieu à du temps de réécriture. Le temps d'écriture effectif a généralement été noté pour chaque texte.

<sup>5</sup> L'insertion de ces syntagmes nominaux démonstratifs a fait l'objet d'un mémoire de Master LICOCO de l'Université Toulouse Jean Jaurès en 2020-2021, celui de Mathilde Forgerit et celle du syntagme nominal démonstratif *ce grand bruit* a fait l'objet d'une communication au CMLF 2020. Cf en bibliographie Garcia-Debanc 1, 2020.

<sup>6</sup> Dans les expérimentations en psycholinguistique, les enfants ont à écrire à partir d'une image, afin d'éviter un amorçage linguistique. C'est le choix qu'ont fait également les chercheurs du LIDILEM pour constituer leur corpus longitudinal.

la compréhension des textes littéraires (Bishop & Boiron, 2018) que pour l'écriture de récits. Reuter (1988) définit trois axes pour le traitement du personnage dans l'enseignement : (a) le personnage comme marqueur typologique, dans la mesure où sa présence est un marqueur du narratif et sa nature un marqueur de genre (les personnages de contes ne sont pas les mêmes que ceux des nouvelles fantastiques ni des romans de science-fiction), (b) le personnage comme élément central de la structure narrative et actancielle et (c) le personnage comme lieu d'investissement du sujet (identification du personnage, empathie...). Les didacticiens ont proposé des activités d'enseignement pour aider des élèves d'école primaire et de collège à donner une existence textuelle aux personnages qu'ils inventent. Des travaux fondateurs ont ainsi mis en évidence les difficultés d'élèves de CM1, CM2 et 6<sup>e</sup> à gérer les risques de confusion entre personnages dans une tâche rédactionnelle (Charolles, 1988) et proposé des activités d'enseignement sur « le mode de désignation des personnages » (Masseron & Schnedecker, 1988 ; Schnedecker, 1995 ; Elalouf, 1999).

Dans la consigne Résolco, la présence des pronoms personnels *elle* et *il* et du syntagme nominal *les enfants* induit l'insertion d'au moins deux personnages de sexes différents et suscite des interrogations sur le statut du SN *les enfants*. Si la consigne impose l'intégration de *elle* et *il* dans une continuité référentielle, en revanche le rédacteur est libre d'attribuer à la dernière phrase une valeur de moralité reposant sur une interprétation générique de *les enfants* ou de décider que ce syntagme renvoie à un référent collectif réunissant les personnages associés à *elle* et *il*. La consigne laisse également libre le choix des personnages, lui-même marqueur du genre narratif choisi : la sorcière et le loup du conte, la vieille femme inquiétante et le fantôme de la nouvelle fantastique ou encore le groupe d'amis du récit réaliste.

Une analyse des configurations de formes linguistiques observées dans les textes rédigés en réponse à la consigne Résolco permettra d'identifier des profils et, à terme, de proposer une cartographie de ce qui peut être raisonnablement attendu à un niveau scolaire donné.

### **3. ANNOTER LA CONTINUITÉ RÉFÉRENTIELLE DANS DES TEXTES D'ÉLÈVES**

Notre point de départ, tel que nous l'avons formulé dans (Garcia-Debanc *et al.*, 2017), est que l'annotation des chaînes de référence dans les textes d'élèves doit pouvoir bénéficier de l'expérience d'annotation conduite dans le cadre du projet Annodis (Péry-Woodley *et al.*, 2011). Cette hypothèse s'appuie sur le fait que les deux projets partagent un même objectif, à savoir l'étude des indices de cohésion référentielle. Dans (Garcia-Debanc *et al.*, 2017), était apparue toutefois nécessaire une adaptation aux spécificités des textes récoltés avec la consigne Résolco, dans la mesure où ces textes narratifs courts d'élèves diffèrent des textes longs publiés traités dans Annodis. La présente étude vise à identifier les formes linguistiques utilisées par les élèves pour construire la cohésion référentielle, ce que nous proposons d'appeler, après Schnedecker (1995), la « continuité référentielle », tandis que l'annotation de textes longs dans Annodis avait conduit à repérer des « chaînes topicales » (Federzoni *et al.*, 2020), c'est-à-dire des segments placés en position de topique dans des zones référentiellement denses.

### 3.1. Des chaînes topicales aux continuités référentielles

Nous reprenons certaines procédures d'annotation des chaînes topicales mises au point dans l'expérience d'annotation Annodis. Cette méthode présente en effet l'intérêt de rendre possible l'annotation d'un phénomène linguistique complexe à grande échelle (dans Annodis, 588 chaînes topicales ont été annotées dans 87 textes, soit 666 000 mots). Elle permet une diversité d'analyses, comme par exemple l'observation exhaustive et la quantification des types de chaînage, du nombre de maillons<sup>7</sup> et des cohabitations entre chaînes (Federzoni *et al*, 2020).

Au-delà du projet Annodis, notre réflexion s'est également nourrie de l'annotation des chaînes de référence proposée dans le cadre du projet Democrat (Landragin, 2015). Ce projet fournit un large corpus diversifié dans lequel sont annotées toutes les expressions référentielles, qu'elles réfèrent à un lieu, une date ou un personnage et qu'elles soient maintenues ou non dans des chaînes de référence (dans Democrat, 20 410 chaînes de référence ont été annotées dans 58 textes, soit 689 000 mots). Les travaux issus de ce projet ont conduit à définir une variété d'indicateurs permettant de caractériser les chaînes de référence. Parmi ces indicateurs, on compte : le nombre de maillons ; la moyenne de la distance entre deux maillons consécutifs ; la catégorie grammaticale et la fonction syntaxique des maillons ; le coefficient de stabilité, qui évalue la stabilité lexicale à l'intérieur d'une chaîne (Schnedecker & Landragin, 2014).

Bien que la notion de chaîne de référence ait pour objectif, comme l'écrit Schnedecker (2019), d'« appréhender le déroulement ou encore le suivi de l'expression référentielle dans sa continuité textuelle ou discursive », la définition qui en est donnée dans la littérature nous est apparue trop contrainte quand il s'agit d'annoter des textes du corpus Résolco, comme nous l'illustrons dans la section suivante. C'est pourquoi nous utilisons la notion de continuité référentielle qui s'inspire des travaux de Givón (1983). Cette notion nous paraît plus adaptée, dans la mesure où les phénomènes en jeu ne renvoient pas toujours strictement aux phénomènes de coréférence et/ou d'anaphore qui caractérisent les chaînes de référence mais permet également de prendre en compte, comme nous allons le voir dans le paragraphe suivant, des éléments qui ne relèvent pas d'une chaîne de référence au sens strict.

Si nous nous contentions d'annoter les chaînes de référence au sens strict, c'est-à-dire les expressions linguistiques qui réfèrent à une même entité, nous laisserions de côté des continuités référentielles qui s'établissent pourtant entre les référents. Ainsi, dans l'exemple de la Figure 1, texte rédigé par un élève de 6<sup>e</sup>, le rédacteur a tissé des liens de continuité référentielle entre les deux personnages associés aux pronoms *elle* et *il* en utilisant alternativement une référence stricte à chacun des deux personnages (*une jeune fille orpheline, Hectorine, elle, un jeune homme nommé Quentin, lui...*) et une référence collective incluant les deux personnages (*ensemble, ils*). Les formes utilisées pour y référer ont été annotées et sont ici délimitées par des boîtes de couleur (bleu pour *Elle* et vert pour *il*). Les liens de continuités sont

---

7 Le terme de *maillon* est utilisé pour désigner les différentes mentions d'une chaîne de référence.

matérialisés par des traits reliant ces boîtes (pour plus de lisibilité, seuls les liens entre les maillons *Elle* sont représentés dans la Figure 1).

Figure 1 : Texte rédigé par un élève de 6<sup>e</sup>, normalisé orthographiquement, dans lequel ont été annotés les expressions référant aux personnages *Elle* et *Il* ainsi que les liens de continuité référentielle concernant le personnage *Elle*

Il était une fois **une jeune fille orpheline qui s'appelait Hectorine**, **Elle** rencontra **un jeune homme nommé Quentin**. **Ensemble, ils** cherchaient des endroits pour s'abriter la nuit, dans la ville. Un jour, **ils** ~~décidèrent de se séparer~~. ~~Quelques années plus tard~~, **elle** habitait dans cette maison depuis longtemps. Une demeure lugubre dans la forêt mais toujours en quête de **Quentin, son ami d'enfance**. Par contre **lui** **Quentin, terrorisé par les histoires de ses hôtes** habitait dans une grotte. Tout à coup, un astéroïde percuta la montagne d'à côté. **Il** se retourna en entendant ce grand bruit. **Lui** et **Hectorine** ~~moururent et se retrouvèrent~~. **Impressionnés, ils** virent un énorme cratère. Les débris de l'astéroïde, rougeoyants, se levèrent et un tremblement de terre surgit. **Quentin** se fit écraser par les rochers dévalant la pente. **Hectorine, horrifiée par cette boucherie** se fit atomiser par des rayons lasers sortant des vaisseaux spatiaux arrivant d'un trou interdimensionnel. Depuis cette aventure, les enfants ne sortent plus la nuit.

FIN

Dans ce texte, le syntagme *les enfants* de la dernière bandelette (P3) n'a pas été rattaché aux personnages de l'histoire et ne peut pas être interprété de manière spécifique. Il ne peut donc être interprété que de manière générique, dans la mesure où, d'une part, les deux personnages ne désignent pas des enfants mais de jeunes adultes, et où, d'autre part, à la fin de l'aventure imaginée, les deux personnages sont morts, respectivement *écrasé par les rochers* pour *il* et *atomisée par des rayons lasers* pour *elle*.

### 3.2. Choix méthodologiques

En premier lieu, nous avons décidé de circonscrire l'annotation des continuités référentielles aux seuls référents présents dans l'ensemble des textes du corpus, autrement dit aux référents provoqués par la consigne Résolco (il est en effet tout à fait possible que les rédacteurs introduisent dans leur récit une diversité d'autres référents, en particulier d'autres personnages). Parmi les référents provoqués par la consigne, seuls les personnages qui assurent la fonction sujet dans les trois phrases, qu'il s'agisse de référents humains ou non-humains, seront donc annotés. En circonscrivant l'annotation aux personnages principaux, l'annotateur est amené à se concentrer sur les référents qui jouent un rôle central dans le récit. Ces référents sont plus propices à donner lieu à des reprises qui « tissent des liens » entre une grande partie des propositions qui composent le texte, contribuant ainsi à construire ce que Halliday & Hasan (1976) appellent la « texture ».

En second lieu, et comme nous l'avons justifié plus haut, nous avons retenu la notion de continuité référentielle plutôt que de coréférence pour dénommer les relations qui unissent les expressions référentielles utilisées dans les textes pour

désigner les référents animés de la consigne Résolco. Ainsi, dans l'exemple donné dans la Figure 1, nous identifions une continuité couvrant l'ensemble du texte et reliant entre elles des expressions renvoyant aux individus et au groupe : *une jeune fille orpheline* > *qui* > *Hectorine* > *Elle* > *Ensemble, ils\** > *ils\** > *elle* > *son* > *Hectorine* > *se retrouvèrent\** > *impressionnés, ils\** > *Hectorine, horrifiée par cette boucherie*. Les astérisques signalent ici les expressions qui renvoient au groupe.

Notre méthode permet d'explicitier, pour chaque expression, le fait qu'elle renvoie à l'individu ou au groupe. Nous retrouvons ici une parenté avec le choix de « modéliser de manière exhaustive les transitions référentielles d'un individu au groupe auquel il appartient et inversement », adopté dans le projet MC4 (Landragin, 2011), choix ensuite abandonné dans le projet Democrat pour simplifier le modèle et permettre une annotation à grande échelle.

#### **4. PREMIÈRE ESQUISSE DES STRATÉGIES DE GESTION DES CONTINUITÉS RÉFÉRENTIELLES**

Une fois l'annotation réalisée, nous disposerons de données dont l'analyse permettra de dresser une cartographie des formes linguistiques utilisées par les rédacteurs pour gérer les continuités référentielles concernant les personnages principaux mentionnés de façon anaphorique dans les phrases imposées de la consigne Résolco. L'exploitation des annotations permettra également d'étudier les stratégies utilisées par les rédacteurs pour introduire, maintenir et faire cohabiter les différents référents dans le discours. En prenant appui sur les travaux consacrés aux processus rédactionnels (Hayes et Flower, 1980 ; Bereiter et Scardamalia, 1987), deux types de manières de faire peuvent d'ores et déjà être identifiées : la stratégie « pas à pas » et la stratégie de « planification ».

##### **4.1. Stratégie « pas à pas »**

La stratégie « pas à pas » consiste à régler l'une après l'autre les questions de référence que pose l'insertion des phrases figurant sur les bandelettes. Le rédacteur qui opte pour cette stratégie procède donc d'abord à l'introduction du référent du pronom *elle* en faisant précéder le collage de la première bandelette (P1) de l'écriture d'une phrase contenant une expression référentielle pouvant servir d'antécédent à ce pronom personnel singulier de genre féminin. Une fois P1 insérée, le rédacteur fait de même pour *il* puis pour *les enfants*. Dans sa version la plus minimale, cette stratégie aboutit à un texte dans lequel l'apparition du personnage renvoyant à *il* entraîne la disparition du personnage référant à *elle*. La Figure 2 fournit un exemple illustrant une stratégie de ce type. Dans ce texte collecté dans une classe de 6<sup>e</sup>, chaque personnage est introduit dans la phrase précédant chaque bandelette et maintenu dans le discours jusqu'à l'apparition du personnage mentionné dans la bandelette suivante.

Figure 2 : Texte rédigé par un élève de 6<sup>e</sup>, normalisé orthographiquement, illustrant une stratégie « pas à pas » d'introduction et de maintien des personnages

Un jour il y a **une vieille dame** ~~qui voulait déménager.~~ **Elle** habitait dans cette maison depuis longtemps.

**Elle** mit **sa** maison en vente. Un jeune couple passa et l'aperçut ils décident d'emménager. La première journée tout se passa bien mais la première nuit des bruits retentissent **Luke le petit** ~~se leva et alla~~ dans le salon puis

**il** se retourna en entendant ce grand bruit.

Et **il** ~~retourna dans son~~ lit en courant. Les autres nuits allaient mieux et maintenant.

Depuis cette aventure, **les enfants** ne sortent plus la nuit.

Ici, le lecteur doit effectuer un calcul inférentiel important pour comprendre que *Luke le petit* est un enfant de la famille nouvelle occupante de la maison, mentionnée par le SN indéfini *un jeune couple*. Le lecteur pourrait aussi être amené à considérer que la mention *Luke le petit* implique que le jeune couple a d'autres enfants. Cependant, cette interprétation nécessite beaucoup d'inférences et de coopération de la part du lecteur. *Les enfants* peut aussi être considéré comme une mention suspendue sans antécédent, dans la mesure où P3 n'est pas précédée de l'introduction d'un référent accessible pour *les enfants*. En l'absence de l'introduction de ce référent, c'est l'interprétation générique du SN pluriel qui est sélectionnée, comme pour le texte de la Figure 1.

D'autres problèmes liés à l'intégration de P3 sont observables dans ce texte. On peut ainsi noter une contradiction entre la moralité finale et le récit : si *la nuit* est introduit dans la phrase précédant P2 (*mais la première nuit des bruits retentissent*) et repris dans la phrase qui précède P3 (*mais les autres nuits allaient mieux*), aucun élément du récit ne permet de justifier le verbe *sortent* de P3.

#### 4.2. Stratégie de « planification »

La stratégie de « planification » consiste à anticiper l'intégration des trois bandelettes en instanciant les référents *elle* et *il* et *les enfants* dès le début du texte. Il s'agit, dans ce cas, pour le rédacteur, de concevoir une histoire qui, dès le départ, implique plusieurs personnages incluant *a minima* un personnage féminin, un personnage masculin et des enfants, qui peuvent être le personnage féminin et le personnage masculin introduits pour instancier les pronoms *il* et *elle*. Cette stratégie requiert la gestion simultanée de plusieurs continuités référentielles et donc une organisation du discours exigeant de faire cohabiter en même temps dans le monde du texte en construction plusieurs référents, de genre et de nombre différents. Cette stratégie peut se réaliser de manière plus ou moins complexe dans les textes des élèves. La Figure 1 fournit un exemple de ce type d'organisation, dans la mesure où les deux personnages associés à *elle* (*Hectorine*) et à *il* (*Quentin*) sont introduits avant P1 et continuent tous deux à exister jusqu'à la fin de l'histoire.

La Figure 3 ci-dessous propose un autre exemple de planification dans lequel les deux personnages référant à *elle* et *il* sont introduits dès le début du texte et maintenus jusqu'à la fin. Contrairement à ce qu'on peut voir dans le texte de la

Figure 1, l'insertion de P3 a été ici anticipée : le référent de *les enfants* est introduit dès les premiers mots du texte par le biais du SN coordonné *Camille et Manu* – la présence du verbe *jouaient* permettant d'inférer que ce sont des enfants.

Pendant, cette stratégie de planification peut entraîner certaines difficultés, notamment si, dans la suite du texte, le rédacteur veut différencier les actions des deux personnages. Dans la mesure où l'entité plurielle est une entité coordonnée, la coordination en fait une entité à part entière, dont les membres sont alors peu accessibles. La reprise pronominale de l'un ou l'autre de ces personnages nécessite donc une activation du référent individu indépendamment du groupe *Camille et Manu*.

Figure 3 : Texte rédigé par un élève de CM1, normalisé orthographiquement, illustrant une stratégie de « planification » où les référents de chaque bandelette sont introduits conjointement dès le début du texte avec transition des chaînes référentielles des individus à une chaîne référentielle de groupe

Dans cet exemple, les trois continuités référentielles annotées et matérialisées par les traits de couleur relient les expressions référentielles suivantes :

- *Camille > elle > je > ma > Elle > Ils\* > ils\* > ils\* > ils\* > eux\* > les enfants\**
- *Manu > Viens > Manu > te > il > Ils\* > ils\* > ils\* > ils\* > eux\* > les enfants\**
- *Camille et Manu > Ils > ils > ils > ils > eux > les enfants*

Notre modèle d'annotation met en évidence ici une transition référentielle entre les chaînes *elle* et *il* et la chaîne *les enfants*, qui commence au milieu du texte.

## 5. PREMIERS RÉSULTATS OBTENUS SUR UN ÉCHANTILLON DE TEXTES

### 5.1. Nombre de maillons pour chacune des continuités référentielles *il*, *elle*, *les enfants*

L'analyse quantitative d'un échantillon d'une centaine de textes (103 textes, soit environ 20 000 mots), répartis selon les trois niveaux scolaires (CE2, troisième année de l'école primaire, 6<sup>e</sup> et 3<sup>e</sup>, début et fin du collège) permet de faire une première comparaison entre les continuités référentielles observées dans les textes des trois niveaux scolaires retenus.

Ne sont pas considérés dans les calculs les textes dans lesquels la consigne Résolco n'a pas été respectée, autrement dit les textes dans lesquels les trois bandelettes n'ont pas toutes été intégrées (soit au total 4 textes de 6<sup>e</sup> et 7 de 3<sup>e</sup>). En revanche, à partir du moment où les bandelettes sont toutes présentes, les textes sont pris en considération dans les calculs et cela même si la mention des personnages ne

figure que dans les bandelettes (ce qui correspond pour *elle* à 3 textes de CE2 ; pour *il*, à 1 texte de CE2, 3 textes de sixième et 4 textes de 3<sup>e</sup> ; pour *les enfants*, à 4 textes de CE 2, 10 textes de sixième, 14 textes de 3<sup>e</sup>). *Les enfants* pouvant être interprété avec une valeur anaphorique ou une valeur générique, des annotations complémentaires seront nécessaires pour apprécier le degré de cohésion entre la moralité finale et le récit antérieur. N'ayant pas accès à de telles annotations complémentaires à l'heure actuelle, pour ces premières analyses, nous avons décidé de traiter globalement les maillons *Les enfants* sans distinguer valeur anaphorique et valeur générique.

Le premier critère de comparaison entre les trois niveaux scolaires repose sur le comptage du nombre moyen de maillons qui composent chaque continuité référentielle incluant les personnages de P1, P2 et P3, à savoir *elle*, *il* et *les enfants*.

Tableau 1 : Longueur moyenne des continuités annotées en nombre de maillons

	nb textes	nb mots	maillons <i>elle</i>			maillons <i>il</i>			maillons <i>les enfants</i>		
			moy	min	max	moy	min	max	moy	min	max
CE2	21	1775	8	1	22	8	1	19	4	1	8
6e	41	6100	13	2	33	11	1	28	7	1	29
3e	41	11808	18	2	57	17	1	63	10	1	59

La comparaison des moyennes du nombre de maillons fait apparaître, comme attendu, une augmentation de ce nombre selon le niveau scolaire : on observe presque un doublement du CE2 à la 6<sup>e</sup> et plus d'un doublement du CE2 à la 3<sup>e</sup>. Cependant, nous avons noté de très grands écarts entre les textes d'un même niveau scolaire, écarts qui s'accroissent avec le niveau : quel que soit le niveau, on trouve des textes composés d'un nombre de maillons extrêmement faible (minimum de 2) comme des textes contenant par exemple une vingtaine de maillons, en CE2 (max=22 maillons *elle*), ou une soixantaine, en 3<sup>e</sup> (max=63 maillons *il*). Ces résultats sur la très grande hétérogénéité des compétences pour un même niveau scolaire confirment des études antérieures sur d'autres corpus (Garcia-Deban, 2010).

Ces variations vont de pair avec la longueur des textes qui passe d'une moyenne de 85 mots en CE2 à 149 mots en 6<sup>e</sup> pour atteindre 288 mots en 3<sup>e</sup>. Sur cette dimension, on observe également des écarts entre textes d'un même niveau : en CE2, la taille des textes varie du simple au triple et en 3<sup>e</sup>, le texte le plus court compte 91 mots alors que le plus long en compte 900.

Les nombres de maillons pour *elle* et pour *il*, dans l'ensemble des textes, sont à peu près équivalents. Concernant *les enfants*, un contraste net se dessine entre d'un côté les textes de CE2 dans lesquels le nombre de maillons pour *les enfants* ne dépasse pas un maximum de 8 maillons, et de l'autre, les textes de 6<sup>e</sup> et de 3<sup>e</sup> qui montrent jusqu'à 29 et même 59 maillons *les enfants*. Cet écart peut être interprété comme une difficulté, pour les élèves de CE2, à articuler cette dernière phrase à l'ensemble du texte précédent, comme y invite l'anaphore résomptive *depuis cette aventure* et, de ce fait, à une carence dans la planification d'ensemble du texte. On

note tout de même qu'un tiers des textes de CE2 incluent *les enfants* dans une chaîne d'au moins 5 maillons.

## 5.2. La place de la première mention du référent de *elle*, *il* et *les enfants* comme indice des stratégies rédactionnelles dominantes

Nous allons maintenant considérer la place dans le texte du premier maillon de la chaîne référentielle incluant *elle* et *il* et interpréter cette configuration comme un indice de la stratégie rédactionnelle dominante.

Le personnage *Elle* peut être introduit avant la première bandelette et le personnage *Il* entre la première et la deuxième bandelette (e.g. ex 2) sous la forme d'une configuration [Elle<P1<Il<P2]. La présence de cette configuration peut permettre d'inférer une stratégie pas à pas dans laquelle l'élève instancie un personnage susceptible d'être repris par le pronom *elle* dans P1. Lorsque l'élève abandonne ce personnage pour introduire un candidat à une reprise par *il* dans P2, il se trouve en situation de gérer chaque référent-personnage l'un après l'autre.

Les deux personnages principaux peuvent aussi être introduits dès le début de l'histoire (ou du moins avant de coller P1), sous la forme de la configuration [Elle&Il<P1]. Ce choix peut être interprété comme manifestation d'une stratégie de planification de la part de l'élève, après prise en compte de l'ensemble des contraintes posées dans la consigne.

Le tableau 2 fournit les résultats de la présence de ces deux configurations.

Tableau 2 : Répartition des textes annotés selon les niveaux scolaires et en fonction du point d'insertion de la première mention du référent de *elle* et de *il*

	nb textes	[Elle<P1<Il<P2]		[Elle&Il<P1]	
		nb textes	%	nb textes	%
CE2	21	7	33	6	29
6 <sup>e</sup>	41	18	44	11	27
3 <sup>e</sup>	41	10	24	16	39
Total	103	35	34	33	32

Dans l'échantillon analysé, la configuration [Elle<P1<Il<P2] est observée dans un plus grand nombre de textes de 6<sup>e</sup> (44%) que de 3<sup>e</sup> (24%). La configuration [Elle&Il<P1] est plus souvent observée dans les textes de 3<sup>e</sup> (39%) que dans ceux de 6<sup>e</sup> (27%). Ces premiers résultats vont dans le sens d'une préférence des élèves de 6<sup>e</sup> pour une stratégie pas à pas, au contraire des 3<sup>e</sup> qui privilégieraient une stratégie de planification, ce qui correspond aux résultats attendus compte tenu des travaux psycholinguistiques sur les processus rédactionnels.

En revanche, les résultats concernant les CE2 ne permettent pas, au vu du faible échantillon de textes analysés, de dégager une stratégie préférentielle. En effet, pour le niveau CE2, les deux configurations retenues ne rendent compte que de 62% des textes. Il faut souligner que ces deux configurations impliquent nécessairement que *Elle* soit instanciée avant P1. Or, dans 38% de textes de CE2, la première mention du personnage *elle* est celle exprimée dans la bandelette (qu'on peut représenter ainsi : [Elle=P1]). C'est également le cas, mais dans une moindre mesure, dans 20% des textes de 6<sup>e</sup> et 22% des textes de 3<sup>e</sup>.

La configuration [Elle=P1] pourrait être interprétée comme un échec dans la tâche de résolution de problèmes de cohésion textuelle proposée. Cependant, il faut être prudent dans l'interprétation des résultats quantitatifs, dans la mesure où P1 peut être placée en initiale du texte, sans que soit instancié antérieurement le personnage *elle*, pour en constituer l'incipit, ce qui est fréquent dans un roman. Cette solution est d'ailleurs fréquemment observée dans des textes réalisés par des rédacteurs plus experts, comme dans le début de texte écrit par un étudiant de Master<sup>8</sup> reproduit ci-dessous :

**Elle** habitait dans cette maison depuis longtemps. **Elle** en avait assez de ce vieux crépi, de ce carrelage craquelé et de ces tentures sales.

**Juliette** l'avait décidé : ce soir, **elle** fugerait. [...]

Le fait que cette dernière configuration soit beaucoup moins fréquemment observée dans les textes de 6<sup>e</sup> et de 3<sup>e</sup> semble indiquer que les élèves de collège ont mieux compris la norme scolaire et disposent d'une gamme de moyens linguistiques plus étendue que les élèves de CE2. Cette interprétation appelle toutefois un nécessaire retour aux textes pour vérifier l'intégration réussie des autres éléments à insérer. Elle met en évidence la nécessité de croiser approche quantitative et approche qualitative.

Divers cas de figure observés dans la succession de l'insertion des différents personnages conduisent à interpréter les observations en termes de continuum entre planification et gestion pas à pas.

## 6. CONCLUSION

Nous avons présenté dans cet article un modèle d'annotation adapté aux textes d'élèves, dont le but est de permettre une analyse systématique et à large échelle des formes linguistiques employées par des rédacteurs en cours d'apprentissage pour créer et maintenir des continuités référentielles dans des textes narratifs en réponse à une tâche problème. Nous ne disposons encore que d'un petit nombre de textes annotés. Les analyses tirant partie de cette annotation serviront de base pour établir une cartographie de l'emploi de ces formes linguistiques dans des textes de niveaux scolaires contrastés.

Les configurations de marques linguistiques observées permettent d'inférer les stratégies mises en jeu par les rédacteurs pour insérer les phrases imposées et résoudre ainsi les problèmes de cohésion textuelle qu'elles suscitent. Nous avons défini des degrés d'expertises différentes, depuis la non résolution des problèmes de cohésion lorsque la seule mention d'un personnage est celle qui figure dans l'une des phrases imposées jusqu'à une planification d'ensemble manifestant l'anticipation de tous les éléments à insérer, en passant par une stratégie pas à pas par résolution locale des problèmes de cohésion. Diverses configurations intermédiaires sont possibles. De premières indications sont données par la distance entre l'antécédent linguistique choisi par le rédacteur et l'élément anaphorique de la

---

8 La tâche a été proposée à des étudiants de Master LiCoCo, Master recherche de Sciences du langage, de Master MEEF Professeur des Écoles et de Master Métiers de l'écriture de l'Université Toulouse Jean Jaurès, et de plusieurs masters en langues, lettres et linguistique de l'Université Catholique de Louvain-la-Neuve dans le cadre d'un cours commun intitulé "la linguistique du texte".

phrase de la consigne. Moins cette distance est grande, plus on peut inférer une résolution pas à pas. Le nombre de maillons de chaque continuité référentielle et la visualisation de la continuité référentielle pour chacun des personnages principaux fournissent également des indices précieux sur l'épaisseur donnée aux personnages et sur leur gestion conjointe ainsi que sur la prise en compte de la mémoire discursive (Béguélin, 1988) du lecteur potentiel.

Ces données pourront être comparées à celles de textes de genres différents et de degrés d'expertise différents.

Nous avons vu que la consigne peut aussi induire un traitement conjoint de ces deux personnages. On peut alors observer si le rédacteur a conscience de la nécessité de réinstancier l'un des deux personnages par un nom propre ou un SN dans la suite du texte, s'il en a besoin, afin de le rendre saillant pour le futur lecteur. La présence d'un tel élément indique une décentration qui anticipe les éventuels problèmes de compréhension du lecteur. Une étude des traces de réécriture et des erreurs orthographiques également annotées et interrogeables dans le corpus RésolCo (ratures, insertions, annotations en marges du texte, erreurs d'orthographe lexicale ou grammaticale) pourrait permettre d'identifier des hésitations quant aux moyens linguistiques à utiliser pour rendre saillant correctement l'un des deux personnages. Ces annotations constituent de précieuses traces du processus de textualisation et de recherche de solutions, qui demanderont à être interprétées. En revanche, l'absence de ratures et les flottements observés dans le nombre grammatical au moment de l'insertion de P2, *il* de P2 enchainant avec un *ils* dans la phrase antérieure, peuvent être considérés comme l'indice d'un traitement pas à pas.

Les premiers résultats issus du traitement d'un échantillon d'une centaine de textes de trois niveaux scolaires différents (CE2, troisième année d'école primaire, 6<sup>e</sup> et troisième) ont fait apparaître des configurations majoritaires pour l'un ou l'autre des niveaux scolaires mais aussi d'importantes variations interindividuelles pour un même niveau scolaire, confirmant les résultats d'études antérieures sur d'autres corpus (Garcia-Debanç, 2010, 2013). Ces premiers résultats demandent à être confirmés sur l'ensemble des données collectées.

## Références bibliographiques

- BÉGUELIN M.-J. (1988), « Alternatives et décisions lexicales dans l'emploi des expressions démonstratives », *Pratiques* 85, 53-87.
- BEREITER C. & SCARDAMALIA M. (1987), *The psychology of written composition*, Hillsdale, Lawrence Erlbaum.
- BISHOP M.-F. & BOIRON V. (2018), « Des usages didactiques du personnage », *Le Français aujourd'hui* 201, 5-11.
- BRISSAUD C. & VIRIOT-GOELDEL C. (éds) (2019), *Enseigner et apprendre l'orthographe aujourd'hui, Repères*.
- CHAROLLES M. (1988), « La gestion des risques de confusion entre personnages dans une tâche rédactionnelle », *Pratiques* 60 (1), 75-97.
- CISLARU G. & OLIVE T. (2018), *Le processus de textualisation. Analyse des unités linguistiques de performance écrite*, Louvain-la-Neuve, De Boeck.
- ELALOUF M.-L. (1999), « Comment les élèves repèrent dans le récit les chaînes de référence et en construisent. Étude descriptive de la fin du cycle 3 à la sixième et propositions didactiques », in J.-P. BENOIT (éd), *L'émergence du discours*, Rennes, CRDP de Bretagne, Collection Lettres ouvertes, 99-127.
- FEDERZONI S. et al. (2020), « Les chaînes topicales dans la ressource ANNODIS », in F. NEVEU, B. HARMEGNIES, L. HRIBA, S. PRÉVOST & A. STEUCKARDT (éds), *7<sup>e</sup> Congrès Mondial de Linguistique Française*, Université de Montpellier 3, France, 6-10 juillet 2020, SHS Web of Conferences, 78 ; DOI: <https://doi.org/10.1051/shsconf/20207811005>
- FORGERIT M. (2021). *Anaphore et interprétation : analyse d'insertion d'expressions anaphoriques dans des récits d'élèves du CE2 à la troisième*, Université Toulouse 2 Jean-Jaurès, Master LiCoCo, mémoire encadré par Myriam Bras et Claudine Garcia-Debanc
- 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* 5, 81-96.
- GARCIA-DEBANC C. (2013), « Marques spatio-temporelles et structuration d'un texte écrit : effets du genre et du niveau d'expertise du rédacteur », in C. Gunnarsson-Largy & E. Auriac-Slusarczyk E. (éds). *Ecriture et réécritures chez les élèves. Un seul corpus, divers genres discursifs et méthodologies d'analyse*, Louvain la Neuve, Académia, collection Sciences du langage. Carrefours et points de vue, 19-39.
- GARCIA-DEBANC C. (2020), « Ecrire et réécrire pour résoudre des problèmes de cohésion textuelle : quel est donc ce grand bruit dans le corpus RESOLCO ? Analyse de récits d'élèves de 9 à 15 ans » in F. Neveu, B. Harmegnies, L. Hriba, S. Prévost & A. Steuckardt (éds), *7<sup>ème</sup> Congrès Mondial de Linguistique Française*, SHS Web of Conferences, 78 ; DOI: <https://doi.org/10.1051/shsconf/20207807021>
- GARCIA-DEBANC C., BONNEMAISON K. (2014), « La gestion de la cohésion textuelle par des élèves de 11-12 ans : réussites et difficultés », in F. Neveu, P. Blumenthal, L. Hriba, A. Gestenberg, J. Meinschaefer & S. Prévost (éds) *4<sup>ème</sup> Congrès Mondial de Linguistique Française*, SHS Web of Conferences, 8 ; DOI: <https://doi.org/10.1051/shsconf/20140801349>
- GARCIA-DEBANC C., HO-DAC L.-M., BRAS M., REBEYROLLE J. (2017), « Vers l'annotation discursive de textes d'élèves », *Corpus* 16, 157-184.
- GEOFFRE T. (2014), « Profils d'acquisition de la morphographie au cycle 3. Vers une caractérisation des parcours des élèves ? », *Repères* 49, 147-168.

- GIVÓN T. (1983), "Topic continuity in discourse: An introduction", in T. Givón (ed), *Topic Continuity in Discourse. A Quantitative Cross Language Study*, Amsterdam, John Benjamins Publishing Company, 1-42.
- HALLIDAY M.A.K. & HASAN R. (1976), *Cohesion in English*, London, Longman group Ltd.
- HAYES J.R. & FLOWER L.S. (1980), "Identifying the organization of writing processes", in L.W. Gregg & E.R. Steinberg (eds), *Cognitive processes in writing*, Hillsdale, New Jersey: LEA, 3-30.
- LANDRAGIN F. (2011), « Une procédure d'analyse et d'annotation des chaînes de coréférence dans des textes écrits », *Corpus* 10, 61-80.
- LANDRAGIN F. (2015), « Description, modélisation et détection automatique des chaînes de référence (DEMOCRAT) », *Bulletin de l'AFIA* 92, 11-15.
- MASSERON C & SCHNEDECKER C. (1988), « Le mode de désignation des personnages », *Pratiques* 60 (1), 98-123.
- MASSERON C. (2005), « Indicateurs langagiers et stratégies scripturales. Du discours à la langue », *Pratiques*, 125 (1), 205-249.
- NONNON É. (2010), « La notion de progression au cœur de tensions de l'activité d'enseignement », *Repères* 41, 5-34.
- PEREZ-BONNEMAISON K. (2018), *Anaphore et référence en production écrite : étude de textes narratifs d'élèves de 9 à 11 ans, du CE2 au CM2*, Thèse en Sciences du langage, ED CLESCO, Université Toulouse 2 Jean Jaurès.
- PÉRY-WOODLEY M.-P., AFANTENOS S., HO-DAC L.-M. & ASHER N. (2011), « La ressource ANNODIS, un corpus enrichi d'annotations discursives », *Traitement Automatique des Langues* 52 (3), 71-101.
- REUTER Y. (1988), « L'importance du personnage », *Pratiques* 60 (1), 3-22.
- SCHNEDECKER C. (1995), « Besoins didactiques en matière de cohésion textuelle : les problèmes de continuité référentielle », *Pratiques* 85, 3-25.
- SCHNEDECKER C. (2019), « De l'intérêt de la notion de chaîne de référence par rapport à celles d'anaphore et de coréférence », *Cahiers de praxématique* [En ligne] 72 ; DOI : <https://doi.org/10.4000/praxématique.5339>.
- SCHNEDECKER C. & LANDRAGIN F. (2014), « Les chaînes de référence : présentation », *Langages* 3, 3-22.
- TAUVERON C. (1995), *Le personnage. Une clef pour la didactique du récit à l'école élémentaire*, Lausanne : Delachaux & Niestlé.